EMERGING TRENDS IN DATA GOVERNANCE

॥न्यायस्तत्र प्रमाणं स्यात्॥
NLU DELHI

Centre for Communication Governance at National Law University Delhi

॥ न्यायस्त्र प्रमाणं स्यात् ॥

NLU DELHI

Published by
National Law University Delhi Press,
Sector 14, Dwarka, New Delhi 110 078

ISBN
978-93-84272-33-3

Patron: Professor (Dr.) Harpreet Kaur
Vice Chancellor (I/c), NLUD
Faculty Advisor, CCG: Dr. Daniel Mathew
Executive Director, CCG: Jhalak M. Kakkar

Supported by
Friedrich Naumann Foundation For Freedom

FRIEDRICH NAUMANN FOUNDATION For Freedom.
South Asia

Edited by
Swati Punia, Shashank Mohan, Jhalak M. Kakkar, and Vrinda Bhandari

Illustrations and Cover design by
Boski Jain

Typeset and Printing by
Naveen Printers, New Delhi

We thank Srishti Joshi and Priyanshi Dixit for editorial assistance. Additionally, we thank our interns and research assistants - Tansi Fotedar, Kaartikay Agarwal, Pranika Goel, and Sanskruti Yagnik for their time and assistance with this publication. Special thanks to the ever-present and ever-patient Suman Negi and Preeti Bhandari for their unending support for all the work we do at CCG.

Emerging Trends in Data Governance

Foreword

India envisions to make this decade a *techade* of opportunities for its people. It is incubating the third largest start-up base in the world as well as the second largest and fastest growing base of digital users/nagriks in the world.

India must recognise the power of empowering its people through effective data governance frameworks to build a resilient economy and lead the way for societies undergoing digital transformation in the Global South. While public discourse and broader legislative and policy objectives underline the idea of people-centric governance and empowering individuals, it is important that the nuts and bolts of the legislative and policy frameworks effectively articulate this vision and enable such implementation. From a practical standpoint, it is important to understand that the success of India's digital transformation story rests on three pillars that are interlinked – Inclusion, Innovation and Implementation - and need to be factored in while designing policies and regulations that govern the use of data. Digital prowess cannot be built on innovation alone. It must be supported by policies and technologies that are inclusive in design and effective in implementation.

In the absence of adequate data governance frameworks and strategy, large swathes of personal and non-personal data churning the wheels of digital transformation are being accumulated and exchanged by both public and private actors to explore new markets and build new products and services. This has led to recurring incidents of data breaches and leaks and cybersecurity incidents. Such a phenomenon can exacerbate and create deep socioeconomic and regional inequalities and a disempowered citizenry. India needs to check this vicious cycle by addressing the legal vacuum in data governance.

While efforts in this direction have been underway for more than a decade, a concrete solution that effectively addresses the rights and responsibilities of its constituent actors in a constitutionally aligned manner remains pending. As India deliberates on the fourth iteration of personal data protection legislation, this edited volume of essays by the Centre for Communication Governance at the National Law University Delhi serves as a timely and excellent resource to analyse trends in data governance and for shaping the broader discourse and horizons of data governance in the context of India as well as Global South.

The guiding thread of this CCG edited volume of essays is framing and operationalising the various elements of empowerment as a compass for creating relationships that are based on trust, transparency and accountability to alleviate power imbalances and asymmetries in the ecosystem. The authors of this CCG edited volume of essays,

draw on domestic and international literature on the intersection of society, law, economics, and technology, to succinctly capture the potential risks to the idea and implementation of Digital India in failing to acknowledge and incorporate the panoply of rights and values accruing to an individual and collectives within the digital ecosystem that flow from the Indian Constitution and jurisprudence of the Indian Supreme Court.

The CCG edited volume of essays entitled '*Emerging Trends in Data Governance*' is an excellent resource to shape the thinking around data governance in line with India's core values of preserving its diversity and culture, by discussing ideas and approaches that help situate the rights and interests of people at the heart of data governance. It engages with a wide range of ontological concepts and models to discuss and deconstruct new terminologies and taxonomy.

By placing people at the heart of data governance, India can lead the way in creating a just and equitable digital society and serve as a model for a range of countries having a diverse socio-cultural fabric like India and who face the challenges of onboarding and protecting the rights and interests of first-time users and vulnerable and marginalised people within their digital landscape. I am confident that this edited volume of essays by CCG will encourage discourse on better modelling of policies for a just and equitable digital society that protects and empowers people online as well as offline.

I congratulate the Centre for Communication Governance for collating this rich volume of essays on data governance and for the immense value their research such as this brings to judges, lawyers, policymakers, industry, students and Indian citizens at large.

Professor (Dr.) Harpreet Kaur
Vice Chancellor (I/c)

About the National Law University Delhi (NLUD)

The National Law University Delhi is one of the leading law universities in the capital city of India. Established in 2008 (by Act. No. 1 of 2009), the University is ranked second in the National Institutional Ranking Framework for the last five years. Dynamic in vision and robust in commitment, the University has shown terrific promise to become a world-class institution in a very short span of time. It follows a mandate to transform and redefine the process of legal education. The primary mission of the University is to create lawyers who will be professionally competent, technically sound and socially relevant, and will not only enter the Bar and the Bench but also be equipped to address the imperatives of the new millennium and uphold the constitutional values. The University aims to evolve and impart comprehensive and interdisciplinary legal education which will promote legal and ethical values, while fostering the rule of law.

The University offers a five year integrated B.A., LL.B (Hons.) and one-year postgraduate masters in law (LL.M), along with professional programs, diploma and certificate courses for both lawyers and non-lawyers. The University has made tremendous contributions to public discourse on law through pedagogy and research. Over the last decade, the University has established many specialised research centres and this includes the Centre for Communication Governance (CCG), Centre for Innovation, Intellectual Property and Competition, Centre for Corporate Law and Governance, Centre for Criminology and Victimology, and Project 39A. The University has made submissions, recommendations, and worked in advisory/consultant capacities with government entities, universities in India and abroad, think tanks, private sector organisations, and international organisations. The University works in collaboration with other international universities on various projects and has established MoU's with several other academic institutions.

About the Centre for Communication Governance (CCG)

The Centre for Communication Governance at the National Law University Delhi (CCG) was established in 2013 to ensure that Indian legal education establishments engage more meaningfully with information technology law and policy and contribute to improved governance and policy making. CCG is the only academic research centre dedicated to undertaking rigorous academic research in India on information technology law and policy in India and in a short span of time has become a leading institution in Asia. Through its academic and policy research, CCG engages meaningfully with policy making in India by participating in public consultations, contributing to parliamentary committees and other consultation groups, and holding seminars, courses and workshops for capacity building of different stakeholders in the technology law and policy domain.

CCG has built an extensive network and works with a range of international academic institutions and policy organisations. These include the United Nations Development Programme, Law Commission of India, NITI Aayog, various Indian government ministries and regulators, International Telecommunications Union, UNGA WSIS, Paris Call, Berkman Klein Center for Internet and Society at Harvard University, the Center for Internet and Society at Stanford University, Columbia University's Global Freedom of Expression and Information Jurisprudence Project, the Hans Bredow Institute at the University of Hamburg, the Programme in Comparative Media Law and Policy at the University of Oxford, the Annenberg School for Communication at the University of Pennsylvania, the Singapore Management University's Centre for AI and Data Governance, and the Tech Policy Design Centre at the Australian National University.

The Centre has had multiple publications over the years including the Hate Speech Report, a book on Privacy and the Indian Supreme Court, and most recently an essay series on Democracy in the Shadow of Big and Emerging Tech. The Centre has launched freely accessible online databases - Privacy Law Library (PLL) and High Court Tracker (HCT) to track privacy jurisprudence across the country and the globe in order to help researchers and other interested stakeholders learn more about privacy regulation and case law. CCG also has an online 'Teaching and Learning Resource' database for sharing research-oriented reading references on information technology law and policy. In recent times, the Centre has also offered courses on AI Law and Policy, Technology and Policy, and first principles of cybersecurity. These databases and courses are designed to help students, professionals, and academicians build capacity and ensure their nuanced engagement with the dynamic space of

existing and emerging technology and cyberspace, their implications for the society, and their regulation. Additionally, CCG organises an annual International Summer School in collaboration with the Hans Bredow Institute and the Faculty of Law at the University of Hamburg in collaboration with the UNESCO Chair on Freedom of Communication at the University of Hamburg, Institute for Technology and Society of Rio de Janeiro (ITS Rio) and the Global Network of Internet and Society Research on contemporary issues of information law and policy.

Table of Contents

Introduction: Emerging Trends in Data Governance

Swati Punia, Jhalak M. Kakkar, and Shashank Mohan[1]

As technology – in particular computing capabilities and data analytics – continues to evolve, and reliance on data processing by both the State and private entities increase, questions around the governance of data become more complex. However, the current scheme of data protection frameworks continues to focus on governing personal data, pivoted in individual rights and consent-based mechanisms. Experience over the last few years of implementing the EU's General Data Protection Regulation has highlighted the inability of personal data protection frameworks to effectively address questions around consent, power imbalances, information asymmetry, and transparency. This has sparked a global debate on shifting the contours of data governance frameworks to include non-personal data within its ambit. As India designs a regulatory framework for data, and the EU strategises under its Digital Decade programme, a paradigm shift is underway. The thinking around data governance is shifting to accommodate a range of rights and obligations and operationalise meaningful consent to engage and empower individuals in the digital economy. It is evident that a focus on just protecting personal data at an individual level is not going to be effective in tackling social asymmetries and imbalances.

Over the last several years, India has discussed various iterations of a personal data protection bill ranging from the version released as part of the Justice Srikrishna Committee Report to the Bill introduced in Parliament in 2019, to the Joint Parliamentary Committee version of the Bill from late 2021 to the recent government draft that has been put out for consultation in 2022. In recent times, the issue of non-personal data has been the focus of various policy frameworks including the Non-Personal Data Governance Framework, the report submitted by the expert committee constituted by the Ministry of Electronics and Information Technology (Meity) under the aegis of Kris Gopalakrishnan on Non-personal Data Governance Framework, Meity's white paper on National Open Digital Ecosystem, and the NITI Aayog's draft framework on Data Empowerment and Protection Architecture (DEPA) released in 2020, to the draft India Data Accessibility and Use Policy in early-2022 and draft National Data Governance Framework Policy in mid-2022.

---

1 Swati Punia, Jhalak M. Kakkar, and Shashank Mohan work at the Centre for Communication Governance at the National Law University Delhi. Their profile can be accessed, along with other editors and authors of this volume, on the page *"Notes on Contributors'* in the end of this volume.

These policy processes have kickstarted conversations in India around the challenges of fluidity between personal data and non-personal data, limitations of the notice and consent mechanism, effecting models of data stewardship and operationalizing group/community data rights while safeguarding individual rights and interests. Ideas of data custodians owing a 'duty of care' to communities and data stewards such as data trusts, data commons, and data cooperatives acting as responsible stewards have sparked dialogue on some of the larger questions of data governance.

This edited volume of essays '*Emerging Trends in Data Governance*' was conceptualised by the Centre for Communication Governance at the National Law University Delhi to prompt the ecosystem to map the way forward on some of the looming questions as well as to question the settled norms. How to situate rights of a group/community alongside an individual's rights? What are the alternative models to consent to build trust and transparency? What are the metrics for separating personal data from non-personal data? How do we calibrate risks attached to different forms of data? How do we make a value assessment for data? What should be the policy approach for testing/embedding new models and approaches in the digital ecosystem? In finding answers to some of these, the edited volume of essays shed light on grey areas, identify trends, gaps, and limitations, and recommend policy approaches and strategies.

The Centre for Communication Governance at the National Law University Delhi has been analysing issues around privacy and data governance closely over the last several years and believes that it is important to deep dive into the various elements and threads within data governance to help ideate existing and emerging concepts and articulate potential approaches. This timely volume has a collection of edited essays written by academics and professionals who are some of the leading thinkers in India on the subject of data governance and privacy. This volume explores ideas and perspectives around – consent as a basis for user empowerment and its intersection with DEPA; the value of data and its drivers; types of data stewardship models and their significance in data governance; group data and related rights of communities; feminist principles and their impact on privacy; and nature of emotional recognition technology and its evolution in the context of data governance. These topics are comprehensively captured in the essays described below:

The volume begins with an essay analysing the effectiveness of DEPA, a tech solution designed to operationalise the concept of data empowerment using the consent model, a contentious central piece to many data protection frameworks in the world. This piece expounds on the limitations and challenges of using consent as a means to empower digital users. The next set of four essays, discuss the idea of empowering individuals by way of collectives in the form of groups or communities. Of these, two essays focus on deconstructing the notion of group or community, their relationship with rights and interests related to data, and the definitional ambiguity enveloping them. The other two essays discuss new approaches and alternate perspectives

for assessing the value attached to data and methods and mechanisms to unlock the value of data that help empower individuals and benefit communities. The idea of data stewardship to unlock the value of data is explored in a subsequent essay in the context of the Indian judiciary. The next essay argues for rejecting the design, development, testing and deployment of emotion recognition technologies. The final essay adds a feminist/gender lens to the ongoing privacy discourse in India.

In the first essay, Smriti Parsheera's piece '***An Analysis of India's New Data Empowerment Architecture***' traces the evolution of DEPA and highlights the fractures in its design and the associated data empowerment narrative. She identifies various challenges in fixating on a consent model for providing real control and agency to users in the digital era. Consent fatigue, poor accessibility and readability of privacy policies, and behavioural, cognitive and structural barriers are some of the gaps discussed by her. She argues that empowerment should not only be about agency over collected data but also having less data available about oneself and that the idea of empowerment needs to be achieved through multiple pathways, and not through singular actions like creating a tech architecture or enacting a law. She suggests looking beyond consent and to consider other models of building accountability, trust and empowerment. She warns against having one particular model endorsed by the State for managing consent as it could stifle the growth of alternate standards and models enabling an effective pathway to empower users.

Arindrajit Basu and Amber Sinha in '***Group Data Rights in Law and Policy***' apply a different lens to empowering individuals in the age of big data and analytics. They sketch out the concept of grouping people as a way to empower communities / groups by offering them some control over the data they generate. While delving into the legal and technical uncertainties around the concept of groups/communities and the data related rights they exercise as a collective versus as individuals (making up the group), they discuss three critical questions: When should data rights vest in a group vis-à-vis individuals of the group? How would group data rights interact with external entities and individuals of the group? And how would the enforcement of group data rights take place for algorithmically determined collectives? They argue that groups may be useful intermediaries for the enforcement of rights, while the rights continue to vest with individuals. They clarify that in case of conflict between an individual's rights and group's rights, the former should prevail. For algorithmically determined collectives, the authors argue for algorithmic transparency as the first step towards understanding how data rights will be exercised in these groups.

Continuing the discussion on group rights over data, Kritika Bhardwaj and Siddharth Peter de Souza argue that the rights of groups or communities must form part of every data regulatory framework in their essay '***Unpacking Community Data – Agency, Rights, and Regulation***'. While examining the discourse on taxonomy and terminologies used for collectives such as a group or a community, they discuss

three concepts: community, community data, and community data governance. They identify the gaps in current policy design, examine definitional challenges, and propose a taxonomy based on identity and interests. While framing the argument on broadening the horizons of the concept of 'community', they propose to look beyond the economic standpoint of extracting commercial value to account for social, cultural, and political identity and interests. They argue that communities could be formed on the basis of needs, work, or social causes and recommend that members of a community should have the agency to understand what and where the data is used, how it is used, and how it impacts them. With regard to governing such entities, they recommend acknowledging the fluid nature of communities, their heterogeneity in terms of how communities are formed, how they operate and disappear, and the ways members of the community associate with data related to them. Incorporating these factors while designing governance frameworks would ensure that the capacity of communities to organise and thrive is not constrained by needless formalisation.

Regarding the design of data governance policies, Mansi Kedia and Gangesh Varma in '***For What it's Worth: Realising the Value of Data***' emphasise the need to incorporate a comprehensive understanding of the different types of values of data. They highlight different dimensions of data that assign different kinds of value to data, namely economic, social and technical. For this purpose, the urge to move the focus beyond the contours of personal data to governing non-personal data as well. They argue for a model of data governance that harnesses the potential of data without causing harm and minimising abuse and misuse. They recommend adopting a multi-stakeholder approach to designing a balanced data governance regime that considers the varying objectives and interests of stakeholders through transparent and participative processes. The essay concludes by emphasising the need to build better institutions and reliable processes that help assess the value of data in a holistic manner for developing effective data governance frameworks. Mansi and Gangesh caution against promoting one set of values alone and disregarding other drivers of the value of data as it would breed inequality and inefficiency in the ecosystem.

Astha Kapoor echoes the need for adopting an alternative approach to data governance in her essay '***Data Stewardship: Re-imagining Data Governance***'. She highlights the insufficiency of the notice and consent regime to protect the privacy of individuals and also argues that existing data protection frameworks do not account for harms that arise at a community level. Astha proposes that the data stewardship model could help address some of these challenges. She discusses the concept of data stewardship to help unlock the value of data in a manner that leads to individuals' empowerment through their active and direct participation in the data economy. Consent in this regard is operationalised through 'data stewards', a class of independent intermediaries obligated to act in the interest of the data generators. She discusses some of the most popular models of data stewardship across the globe

- data cooperatives, data commons, personal data stores, and data trusts - that are attempting to enhance the decision-making powers of individuals and communities to enable them to reap economic benefits from their data. Acknowledging the challenges of actualising this bottom–up approach, she emphasises the need to record and learn from ground-level evidence and deploy sandboxes to support evolution and innovation in this space.

Ameen Jauhar explores the potential of the data stewardship model within the judicial system. In his essay, '***Making Data Count - A Case for Developing Data Stewardship Models for the Indian Judiciary***', he focuses on operationalising the idea of data trusts for non-personal data within the judicial system to achieve the twin objective of accumulating digital intelligence and judicial data sharing for social innovation. Amongst stewardship models such as data exchanges, data cooperatives and data trusts, he finds data trusts as the most suitable stewardship model in the context of the Indian judiciary. He recommends that such a data trust must be independent in nature, should be a non-profit entity, and must incorporate an institutional and technological layer. Moreover, this entity should also have a framework for discharging fiduciary obligations as well as engagement protocols and mechanisms.

Vidushi Marda in '***Emotion Recognition and the Limits of Data Protection***' captures the evolution and nature of emotion recognition technology. She explains that facial expressions are not solely related to emotional states and they have multiple causes and meanings. She draws from other jurisdictions like the EU to make a case for banning this form of technology for its inherent invasive and discriminatory nature and direct impact on human rights and freedoms including dignity, autonomy, privacy, etc. Vidushi explains why data protection frameworks are inadequate to regulate such technologies and the need to employ a slew of regulatory tools and levers to check the use of emotion recognition technology and its impact on society. Vidushi examines in detail why the data protection laws proposed in India as well as in the UK, US, Brazil, and China are inadequate in mitigating the harms arising from emotional recognition technology due to the wide exemptions given to the State. She attributes the growing appetite for its adoption to proliferation of surveillance technology for purposes of public security, national security, and public order across the globe and the trend of affording exemptions in the areas and relationships that require protection is not unique to India. She makes a thought-provoking case for why data protection frameworks are not the site for dealing with the dangers of surveillance, oppression, marginalization and criminalization of communities. According to her, the data protection frameworks are primarily concerned with efficient and safe data processing, instead of challenging the growth and ubiquity of surveillance. Therefore, she recommends that the best way to deal with emotion regulation technologies is to reject the design, development, testing, and deployment of such systems.

In *'Searching for a Room of One's Own in Cyberspace: Datafication and the Global Feministation of Privacy'* Anja Kovacs analyses the modern data economy through the lens of gender, sexuality, and autonomy. She discusses how privacy as a concept has been mobilised and interpreted to fundamentally curtail the decisional autonomy of women and gender and sexual minorities eroding their right to self-determination. She argues that due to rapid datafication, such a predicament now applies to all individuals. She discusses trends that lie at the heart of the global feminisation of privacy: how consent and anonymity are mobilised by companies and governments to drive datafication leading to a reconfiguration of the public and private divide, and the rise of dataveillance that drives it. Anja states that there is a veritable paradigm shift in the conceptualisation of our bodies. She explains the need to evolve our thinking from treating our virtual bodies merely as a reflection of our physical bodies to understanding that our bodies now need to comprehensively incorporate both virtual and physical forms of bodies.

As India formalises its data governance and internet governance frameworks in the coming months, these essays offer a snapshot of the challenges that must be considered and potential approaches that may be adopted as these policy and legislative proposals advance to the next stage of becoming law. We hope these essays help future researchers better understand and analyse India's data governance trajectory. Additionally, we believe that the analysis in these essays on the emerging trends in data governance in India articulate approaches and thinking that may be relevant across the world, particularly in the Global Majority. The world is closely looking at India for the perspectives it embodies, policies and strategies it adopts, and discourses it shapes in the region to alleviate digital harms/ risks and elevate the rights and interests of individuals. Many countries from the Global Majority, in particular those transitioning to the digital economy, are looking towards India for inspiration and guidance on developing legislative and policy frameworks that weave together the local contexts and global realities.

We would like to extend our deep gratitude to all the contributors to the edited volume of essays, the National Law University Delhi and to our partners for supporting us in putting together this timely volume on emerging trends in data governance.

An Analysis of India's New Data Empowerment Architecture

Smriti Parsheera[1]

Introduction

In August 2020, India's official think tank NITI Aayog put out a discussion paper on the Data Empowerment and Protection Architecture (DEPA).[2] DEPA is a technology-enabled architecture that relies on user consent to facilitate personal data sharing through verifiable records. It is the brainchild of the Indian Software Product Industry RoundTable (iSPIRT), a private think tank born out of the Aadhaar project that also supported the NITI Aayog in the preparation of the DEPA discussion paper.[3] DEPA represents the consent layer, the fourth layer, of the India Stack framework – a set of technological solutions that involve the use of application programming interfaces (APIs) to deliver what are often described as digital infrastructure platforms.[4] The other three layers of this Stack, as originally envisaged by iSPRIT, -- the presence-less, paperless, and cashless layers, have come into effect through projects like Aadhaar authentication, DigiLocker and the Unified Payment Interface (UPI).[5]

As the fourth layer of India Stack, the DEPA framework focuses on creating technological means for the organised collection and verifiability of consent, which is seen as the basis for processing one's personal data. The crux of DEPA's technical architecture lies in an electronic consent artifact, which was released by the Ministry of Electronics and Information Technology (MeitY) in 2017.[6] As per MeitY, such an artifact should contain identifying details of the entities permitted to share and receive data, permissible purposes, data types, duration of use, logs for auditing, and a digital signature. Standardised APIs will then facilitate data sharing

---

1 Smriti is currently a Fellow with the CyberBRICS Project at FGV Law School, Brazil and a PhD candidate at IIT Delhi's School of Public Policy. She can be reached at smriti.parsheera@gmail.com.

2 NITI Aayog, 'Data Empowerment And Protection Architecture - Draft for Discussion' <https://www.niti.gov.in/sites/default/files/2020-09/DEPA-Book.pdf> accessed 25 October 2022.

3 ibid.

4 Vivek Raghavan, Sanjay Jain and Pramod Varma, 'India Stack---Digital Infrastructure as Public Good' (2019) 62 Communications of the ACM 76 <https://dl.acm.org/doi/10.1145/3355625> accessed 25 October 2022.

5 Originally envisioned by iSPIRT, these technologies are now owned and operated by different organisations, such as the Unique Identification Authority of India, the Ministry of Electronics and Information Technology and the National Payments Corporation of India. 'India Stack' <https://indiastack.org/faq.html> accessed 26 October 2022. However, over time, the term India Stack has come to signify a broader suite of technical architectures, including projects such as Co-Win, Aarogya Setu, and government e-marketplace (GeM).

6 Ministry of Electronics and Information Technology, 'Electronic Consent Framework - Technology Specifications, Version 1.1' <https://dla.gov.in/sites/default/files/pdf/MeitY-Consent-Tech-Framework%20v1.1.pdf> accessed 25 October 2022.

between information providers and users using the consent artifact with the help of a new category of intermediaries called consent managers.

Over the last few years, several government agencies have endorsed and implemented DEPA in various forms, sometimes by different names, and in a few cases even before the idea of DEPA was officially articulated in 2020. For instance, the consent layer of India Stack, now known as DEPA, has been endorsed in MeitY's technical specifications on electronic consent,[7] the Reserve Bank of India (RBI)'s Account Aggregators framework,[8] and the Ministry of Health and Family Welfare's reliance on consent managers in the Ayushman Bharat Digital Mission.[9] In light of these developments, DEPA has easily transitioned from a thought experiment into a policy reality involving an evolving ecosystem of actors, with the eventual goal being to expand its application across all sectors.

This paper traces the evolution of DEPA and studies its focus on user consent as the means to operationalize data empowerment. The idea of DEPA rests on two foundational pillars of 'data empowerment' and 'consent'. I begin in Section 1 with a discussion on the rise of the data empowerment narrative, ways in which 'empowerment' has been interpreted and DEPA's place in that narrative. This is followed in Section 2 by an explanation of DEPA's institutional structure and current path. Given DEPA's focus on consent as the basis for user empowerment, I turn next in Section 3 to a discussion on the consent conundrum -- how consent in the information age is recognised to be broken for several reasons but still remains an indispensable part of informational privacy frameworks. This discussion is vital to the subsequent analysis of DEPA's effectiveness as a solution to the consent problem. This is covered in Section 4, which offers an analysis of DEPA's positive and negative aspects, both in terms of its design features and broader questions of process and governance. Section 5 contains the concluding remarks.

1. Rise of the Data Empowerment Narrative

The goal of empowerment has come up in several Indian policy documents. It is a key part of the NITI Aayog's discussion paper on DEPA, which is the focus of the present discussion. Empowerment was also mentioned in the title of the data protection report released by the Justice Srikrishna led Committee of Experts, which identified it as one of the ingredients of a free and fair digital economy.[10]

---

7 ibid.

8 Reserve Bank of India, 'Reserve Bank of India, Master Direction- Non-Banking Financial Company - Account Aggregator (Reserve Bank) Directions' (2016) RBI/DNBR/2016-17/46 <https://www.rbi.org.in/Scripts/BS_ViewMasDirections.aspx?id=10598> accessed 26 October 2022.

9 Ministry of Health & Family Welfare, 'National Digital Health Blueprint' (2019) <https://main.mohfw.gov.in/sites/default/files/Final%20NDHB%20report_0.pdf> accessed 25 October 2022.

10 Committee of Experts under the Chairmanship of Justice B.N., 'A Free and Fair Digital Economy: Protecting Privacy, Empowering Indians' (2018) <https://www.meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf> accessed 25 October 2022.

In the Committee's words, '*a free and fair digital economy that empowers the citizen can only grow on the foundation of individual autonomy, working towards maximising the common good*.'[11] Even the stated vision and mission of Aadhaar is to empower residents -- to be able to authenticate their identity anytime, anywhere.[12] But what exactly does empowerment mean in the personal data context?

In a 2014 agenda document, the World Economic Forum highlighted the need to rethink personal data through the lens of trust. It identified transparency, accountability, and empowerment of individuals as the three pillars of trust, defining empowerment as consisting of two elements. First, individuals having a say in how their data is used by organizations. Second, having the capacity to use their data for their own purposes.[13] Another useful framing published by the World Wide Web Foundation explains that individuals should not just be seen as passive beneficiaries of data empowerment but as active agents of change. This implies the ability to be involved in decisions about the collection of data (and not just decisions about subsequent access, use, and sharing). The authors note that '*data is never neutral and those who control production have influence over what gets collected, how it's used, and therefore its outcomes*.'[14] Empowerment in this broader sense would involve a structural shift in existing data models. Data stewardship models, which suggest the creation of data cooperatives and trusts with direct participation and control by the community, are illustrative of this brand of empowerment.[15]

Yet another interpretation of empowerment, which is rooted in the conceptualization of data as an economic resource, is that of seeking a 'fair value exchange' for data. A report by the UK based Citizen Advice Bureau explained this to mean that individuals should be able to get a clear benefit from sharing their personal data. Accordingly, they include the ability to exercise control over the benefits that consumers wish to derive from their data as a component of personal data empowerment.[16] In the Indian context, Nandan Nilekani has often relied on a similar narrative to make a case for 'data rich' Indians being able to extract better value from their data.[17]

---

11 ibid.

12 UIDAI, 'Vision & Mission' <https://uidai.gov.in/en/about-uidai/unique-identification-authority-of-india/vision-mission.html> accessed 25 October 2022.

13 World Economic Forum and A.T. Kearney, 'Rethinking Personal Data: A New Lens for Strengthening Trust' (2014) <https://www.weforum.org/reports/rethinking-personal-data/> accessed 25 October 2022.

14 Andreas Pawelke and Michael Cañares, 'From Extraction to Empowerment: A Better Future for Data for Development' (World Wide Web Foundation, 11 May 2018) <https://webfoundation.org/2018/05/from-extraction-to-empowerment-a-better-future-for-data-for-development/> accessed 25 October 2022.

15 Astha Kapoor, 'Practising Data Stewardship in India, Early Questions' (23 October 2020) <https://www.adalovelaceinstitute.org/blog/practising-data-stewardship-in-india/> accessed 25 October 2022.

16 Liz Coll, 'Personal Data Empowerment: Time for a Fairer Data Deal?' (Citizens Advice Bureau 2015) <https://www.citizensadvice.org.uk/Global/Public/Corporate%20content/Publications/Personal%20data%20empowerment%20report.pdf> accessed 25 October 2022.

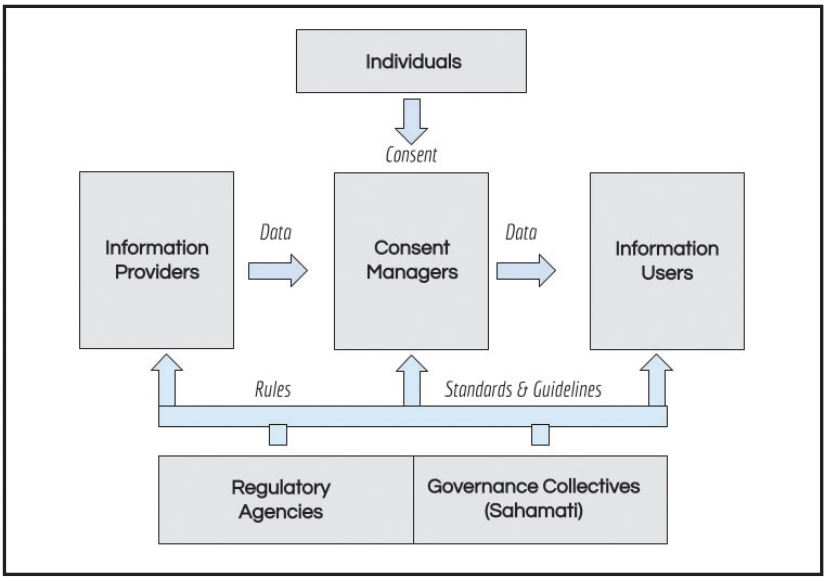
17 Nandan Nilekani, 'Data to the People' <https://www.foreignaffairs.com/articles/asia/2018-08-13/data-people> accessed 25 October 2022; Nandan Nilekani, 'India Can Offer a Radically New Way of Looking at Data: Nandan Nilekani' (*ThePrint*, 9 August 2018) <https://theprint.in/india/governance/india-can-offer-a-new-way-of-looking-at-data-nandan-nilekani/94860/> accessed 25 October 2022.

While there is scope for significant disagreement on whether inherent agency over one's data or interest in extracting its economic value should be the driving factor, all approaches discussed above are consistent in their regard for more user control as a tool for empowerment. In the case of DEPA, which shares its origin with Nilekani's thinking, reliance on user consent and the verifiability of that consent is seen as the source for empowerment. This is based on the premise that in order to be data empowered, individuals should have the practical means to access, control, and selectively share their personal data. Further, the documentation on DEPA also speaks of the empowerment of small and medium enterprises that currently do not have the ability to access their data that is locked in silos.[18]

2. DEPA's Architecture and Adoption

DEPA's functioning relies on the interaction between its technical components – consisting of the electronic consent artifact and API specifications to enable data exchanges – and the institutional arrangements required to facilitate these exchanges. As illustrated in the figure below, there are six main sets of actors involved in the implementation of DEPA.

Figure 1: Actors in the DEPA Ecosystem



Individuals: Individuals whose personal data will be collected and shared through the DEPA ecosystem are supposed to make decisions about what types of data can be shared, with whom and for what duration and purpose. They can transmit their

18 NITI Aayog (n 2).

consent for the selected purposes through the consent artifact, which will be enabled by consent managers.

Consent managers (or Account Aggregators in the financial sector): This refers to a new class of intermediaries that will facilitate the sharing of data on the basis of valid consent from the individual. As per DEPA's formulation, consent managers are supposed to be 'data blind' by design, which means that they can only facilitate encrypted data flows without being able to see the data themselves.[19] The concept of consent managers also found a place in the Personal Data Protection Bill, 2019 (PDP Bill) that has since been withdrawn by the government.[20] The PDP Bill described a consent manager as a type of data fiduciary that would enable individuals 'to gain, withdraw, review and manage' consent using an accessible, transparent and interoperable platform'.[21] The PDP Bill proposed that all consent managers, which would include managers under the DEPA framework, would need to be registered with the proposed Data Protection Authority.

Information providers and users: The entities between whom data sharing can take place based on the consent of the individual. The current design of DEPA provides for a principle of reciprocity, which means that an entity can access data as an information user only if it also agrees to become a data sharer in the system. The list of entities that are at various stages of becoming information providers and users under RBI's Account Aggregators framework, which is DEPA's first application, includes banks, credit companies, and investment advisors.[22]

Regulatory agencies: Table 1 below provides a timeline of key developments surrounding DEPA, as reported in NITI Aayog's discussion paper.[23] It illustrates that the adoption of DEPA is currently taking place on a sector-by-sector basis with government departments and regulatory agencies setting the norms around how consent managers will operate in their domains. In the financial sector, the RBI's announced the Account Aggregators framework in 2016,[24] a limited version

---

19 ibid.

20 The Personal Data Protection Bill, 2019, Lok Sabha (Bill No. 373 of 2019)' <http://164.100.47.4/BillsTexts/LSBillTexts/Asintroduced/373_2019_LS_Eng.pdf>. In August 2022, the government withdrew the PDP Bill indicating the roll out of a new comprehensive version of the bill in the near future. *The Hindu* Bureau, 'Union Government Rolls Back Data Protection Bill' The Hindu (New Delhi, 3 August 2022) <https://www.thehindu.com/news/national/union-government-rolls-back-data-protection-bill/article65721160.ece> accessed 8 November 2022.

21 Explanation to Section 23, The Personal Data Protection Bill, 2019, Lok Sabha (Bill No. 373 of 2019). The report prepared by the Joint Parliamentary Committee has recommended that this should become a standalone definition rather than being an explanation within Section 23. Report of the Joint Committee on the Personal Data Protection Bill, 2019 (16 December 2021), <http://loksabhaph.nic.in/Committee/CommitteeInformation.aspx?comm_code=73&tab=1>.

22 'Sahamati | Certified Entities in the Account Aggregator Ecosystem' (*Sahamati*) <https://sahamati.org.in/certified-entities/> accessed 26 October 2022.

23 NITI Aayog (n 2).

24 Reserve Bank of India, 'Reserve Bank of India, Master Direction- Non-Banking Financial Company - Account Aggregator (Reserve Bank) Directions' (n 8).

of which came into effect in 2021.[25] Consent managers are also recognised to be one of the building blocks for the management of electronic health records under the Ayushman Bharat Digital Mission.[26] Accordingly, the implementation of the DEPA framework in the health sector will be governed by the norms to be set out by the Ministry of Health & Family Welfare and the National Health Authority. Further, the PDP Bill had proposed that the Data Protection Authority would have the authority to frame regulations to specify the technical, operational, financial and other conditions governing consent managers.[27]

Governance collectives: A non-profit body called the DigiSahamati Foundation (Sahamati) has been established to encourage further adoption and to formulate technical standards and codes of conduct for the Account Aggregators ecosystem. Notably, Sahamati will also perform a governance function by monitoring compliance with the guidelines and standards that it frames. As per the details on its website, Sahamati was founded by iSPIRT volunteers, who are also the designers of DEPA.[28]

Table 1: Timeline of Key Developments on DEPA

Date	Development
September 2016	RBI released the policy on Account Aggregators
March 2017	MeitY adopted the Electronic Consent Framework
August 2017	Formal launch of DEPA *(although it already existed in concept as the consent layer of India Stack)*
October 2019	Account Aggregators launched in securities, insurance, and pensions sectors
November 2019	Reserve Bank Information Technology Private Limited (ReBIT) published the Account Aggregator Ecosystem API Specifications
November 2019	Ministry of Health and Family Welfare released the National Digital Health Blueprint identifying consent managers as one of its building blocks

25 As per the information put out by Sahamati, five Account Aggregators have received an operating license and three others have an in-principle approval from the RBI. Further, four of the entities have already launched client-facing apps.'List of Account Aggregators in India | AA Apps in India' (*Sahamati*) <https://sahamati.org.in/account-aggregators-in-india/> accessed 8 November 2022.

26 Ministry of Health & Family Welfare (n 9).

27 Section 23(5), PDP Bill, 2019 (n 19).

28 'Sahamati | Certified Entities in the Account Aggregator Ecosystem' (n 22).

December 2019	Concept of consent manager included in the PDP Bill, 2019. This was not part of the Justice Srikrishna Committee's draft Bill of 2018.
August 2020	TRAI workshop on Telecom Subscriber Empowerment in which the regulator reportedly discussed that telecom companies would be allowed to become information providers in the Account Aggregator system.
August 2020	The GST Department wrote to the RBI to join the Account Aggregator network
August 2020	NITI Aayog released a discussion paper on DEPA
September 2021	RBI's Account Aggregator framework went live

Source: NITI Aayog DEPA Discussion Paper, 2020

3. The Consent Conundrum

The concept of privacy, of which informational or data privacy is a key part, is deeply rooted in the values of liberty and autonomy. Privacy drives the ability of individuals to make choices about their bodies, minds, homes, relationships, and preferences and sets bounds around how others may interact with these spaces.[29] In the personal data context, this translates into the individual's ability to exercise control over who can use their data, in what contexts, and for how long. Equally, what parts of their personal data can be shared further and with whom? In a bid to achieve this, most data protection frameworks outline a set of dos and don'ts for data users and create institutional frameworks to oversee observance with those norms. But a large part of the weight of these frameworks continues to rest on the shoulders of what are referred to as the 'notice and consent' clauses.

For instance, the (now withdrawn) PDP Bill provided that, subject to certain exceptions, personal data should only be processed with the consent of the individual. In order to be valid, such consent would need to be freely given, informed, specific, clear, and capable of being withdrawn.[30] The Bill also identified the basic types of information that a data fiduciary -- a body that collects and processes personal data – would have to convey to the individual in order for the consent to be informed. This includes information on the types of data being collected, its purpose, data sharing arrangements, and likelihood of cross-border transfer.[31] Consent also made

29 'What Is Privacy?' (*Privacy International,* 23 October 2017) <http://privacyinternational.org/explainer/56/what-privacy> accessed 26 October 2022.

30 Section 11, PDP Bill 2019 (n 19).

31 Section 7, PDP Bill 2019 (n 19).

an appearance in several other parts of the Bill, sometimes becoming the basis to override statutory protections within the Bill itself. For example, the PDP Bill restricted the data fiduciaries from retaining any personal data beyond a period that might be necessary to satisfy the purpose for which it was collected. However, this requirement could be overridden with the explicit consent of the individual.[32]

The emphasis on consent in the withdrawn PDP Bill, and in data protection principles generally[33], stems from respect for individual autonomy, and rational limits on the state's interference in private dealings. It is also fueled by assumptions about the existence of the 'privacy pragmatic' user, which suggests that, given the right information, individuals are best placed to make reasoned decisions about the management of their personal data.[34] While this is sound logic, and also in line with the moral underpinnings of privacy as autonomy, there are several barriers to acting as a privacy pragmatic user in the information age.

Understanding the key elements of this 'consent problem' is integral to evaluating DEPA's effectiveness as a proposed solution. It is well acknowledged that the volume of transactions that an average individual enters into has become so large that it is practically impossible for them to actually read and understand the fine print of all personal data dealings.[35] This was humorously illustrated by the UK company Gamestation that managed to get 88% of customers to transfer legal ownership of their soul to the company through a clause embedded in its terms.[36]

Besides volume, poor accessibility and readability of privacy policies also poses a problem. In a study of the privacy policies of five big digital players in India, we found that policies tend to be laden with legal terms that are designed to avoid legal liability rather than conveying meaningful information to the user.[37] Our survey revealed that even English-speaking university students, which included law students, struggled to understand the terms. The net result being that people do not read privacy policies and those who try, may often fail to understand them.

---

32 Section 9, PDP Bill, 2019 (n 19).

33 Regulation (EU) 2016/679 of the European Parliament and of the Council of 7 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] L119/1, art 7; 'The OECD Privacy Framework' (OECD, 2013) <https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf>

34 Ponnurangam Kumaraguru and Lorrie Faith. Cranor, 'Privacy Indexes : A Survey of Westin' s Studies' (Institute for Software Research 2005) Paper 856 <http://repository.cmu.edu/isr/856> accessed 26 October 2022.

35 For instance, on an average, the Indian smartphone user has about 70 apps on her phone. Spending even half an hour reading each policy would translate to about 35 hours of reading time. Smriti Parsheera, 'Notice, Consent, Privacy: Why We Need to Do Better' Hindustan Times (21 August 2019) <https://www.hindustantimes.com/opinion/notice-consent-privacy-why-we-need-to-do-better/story-bIlGQtuKjOCkfQVxwfGHWO.html> accessed 26 October 2022.

36 Joe Martin, 'GameStation: "We Own Your Soul"' (Bit-Gamer, 15 April 2010) <https://bit-tech.net/news/gaming/pc/gamestation-we-own-your-soul/1/> accessed 26 October 2022.

37 Rishab Bailey and others, 'Disclosures in Privacy Policies: Does "Notice and Consent" Work?' [2018] SSRN Electronic Journal <https://www.ssrn.com/abstract=3328289> accessed 26 October 2022.

The effectiveness of consent is also shaped by various behavioural and cognitive limitations and structural barriers.[38] For instance, individuals often struggle to understand their own preferences or the long term harms from the sharing or misuse of their personal data, particularly when such harms are intangible in nature.[39] Very often, they also do not have the visibility to understand, let alone control, the secondary uses of their data or the other sources with which it may be combined. The lack of choice and power asymmetry between the individual and data fiduciaries also complicates the equation. This is reflected, for instance, in the plight of a WhatsApp user who has little choice but to accept the company's onerous terms in order to stay in touch with family and friends on that network. It is equally applicable to an individual who registered on the CoWin app to gain access to a vaccine and ended up being issued a health ID in the process.[40]

This is the consent conundrum of data protection. Consent is recognised to be broken for several reasons, yet it remains an indispensable part of informational privacy frameworks. To be fair, modern data protection laws try to account for this conundrum by including a host of rights and obligations that apply over and above the requirement of obtaining content. But clearly, a lot more needs to be done. Suggestions in this regard have included the need to reimagine consent, for instance by drawing upon feminist perspectives on how consent should be practiced.[41] Others have also proposed a structural shift to look beyond consent at other models of accountability, trust, and empowerment.[42] In the next section I take a closer look at the extent to which the DEPA framework seems to address the consent problem.

4. An Analysis of DEPA: What Works and What Doesn't?

A review of the DEPA discussion paper suggests that the architecture sets out to achieve two main objectives -- i) improving the way in which individuals consent to the collection and use of their data, ii) facilitating data sharing and portability.[43] These are also among the goals of most data protection frameworks, including the withdrawn PDP Bill. The Bill identified consent as the primary basis of data

38 Daniel J. Solove, 'Privacy Self-Management and the Consent Dilemma' (2013) 126 Harvard Law Review 1880.

39 Alessandro Acquisti, Laura Brandimarte and George Loewenstein, 'Privacy and Human Behavior in the Age of Information' (2015) 347 Science 509 <https://www.science.org/doi/10.1126/science.aaa1465> accessed 26 October 2022.

40 Sarthak Dogra, 'Took Covid Vaccine Using Aadhaar? Your National Health ID Has Been Created without Your Permission - India Today' *India Today* (24 May 2021) <https://www.indiatoday.in/technology/features/story/took-covid-vaccine-using-aadhaar-your-national-health-id-has-been-created-without-your-permission-1806470-2021-05-24> accessed 26 October 2022.

41 Anja Kovacs and Tripti Jain, 'Informed Consent—Said Who? A Feminist Perspective on Principles of Consent in the Age of Embodied Data—A Policy Brief' (Internet Democracy Project 2021) <https://internetdemocracy.in/policy/informed-consent-said-who-a-feminist-perspective-on-principles-of-consent-in-the-age-of-embodied-data-a-policy-brief> accessed 26 October 2022.

42 Rahul Matthan, 'Takshashila Discussion Document - Beyond Consent: A New Paradigm for Data Protection' <https://takshashila.org.in/research/discussion-document-beyond-consent-new-paradigm-data-protection> accessed 26 October 2022; World Economic Forum and A.T. Kearney (n 13).

43 NITI Aayog (n 2).

processing and confers data access and portability rights on individuals.[44] It may be fair to assume that the subsequent iteration of the bill will retain these concepts, making it logical that appropriate technical tools and protocols would need to evolve to give effect to the suggested legal requirements. Seen in this light, DEPA contains some useful design elements that could serve as valuable components in the user empowerment toolkit. Notably, the generation of verifiable consent logs adds an important element of accountability to the system and efficient sharing of data at scale necessarily requires the use of APIs.

However, DEPA is not just another technical standard. As its name and trajectory suggest, it is an elaborate technical and organisational architecture that has been privately developed and is being brought into effect with the backing of the state. DEPA has already seen traction in the financial and health sectors and the aim of its developers seems to be to make this the default form of consent management for all data transactions in the public and private sphere — the DEPA discussion document gives examples of social media, e-commerce, education, and employment as some of the other areas for the deployment of consent managers.[45] It also talks about each government department becoming an information provider. Further, in terms of scope, the framework seems to govern not only how consent will be collected but also the model to be followed by all consent managers, applicable standards accompanying consequences for non-compliance. For all these reasons, DEPA is an important development that merits more detailed discussions. A recent working paper by Tripti Jain makes a significant contribution in this direction.[46] Jain analyzes the Account Aggregators framework against six feminist principles of consent articulated in her previous work with Anja Kovacs,[47] using that as the basis to suggest certain regulatory and technical changes in the existing design of the system.

The inputs submitted by stakeholders to NITI Aayog's discussion paper and the general commentary on Account Aggregators also highlight several pros and cons of the DEPA framework. This includes praise for reducing friction in data transfers, following an ecosystem building approach, and high regard for consumer dispute resolution.[48] At the same time, researchers have also pointed to several gaps and

---

44 Sections 11, 17, 19, PDP Bill, 2019 (n 19).

45 NITI Aayog (n 2).

46 Tripti Jain, 'Tech Tools to Facilitate and Manage Consent: Panacea or Predicament? A Feminist Perspective' <https://datagovernance.org/files/research/1641969121.pdf> accessed 26 October 2022.

47 Anja Kovacs and Tripti Jain (n 41).

48 Pramod Rao, 'Dispute Resolution In The Account Aggregator Ecosystem, Anchored By "Sahamati"' (*BQ Prime*) <https://www.bqprime.com/opinion/dispute-resolution-in-the-account-aggregator-ecosystem-anchored-by-sahamati> accessed 26 October 2022.

inadequacies in the framework.[49] This includes its overemphasis on consent, lack of clarity on functions of consent managers, exclusion of non-smartphone users, and potential for overuse of personal data. In the discussion that follows, I add to this analysis with some observations on the challenges posed by the manner in which DEPA is currently taking shape. However, the DEPA discussion paper makes it clear that the project is still a work in progress and is meant to remain evolving and agile in nature. Accordingly, further observations about the project are also likely to evolve based on its outcomes and future directions.

4.1. Implications for Innovation and Competition

The creation of DEPA is premised on the assumption that there is a certain model of data empowerment, based on individual agency and data value extraction, that is most suitable for the people of India. Although most would agree that more agency in the hands of individuals is a desirable goal, there can be many paths to achieving that end. For instance, the Citizen Advice Bureau's report, referred to earlier, offers a detailed mapping of the different types of tools and services, enabling infrastructures, and decision support services that can facilitate personal data empowerment.[50] This includes tools relating to transparency, data access, data storage, permissions management, and building personal profiles, all of which offer different paths to empowerment. DEPA focuses only on a subset of these issues, with an emphasis on consent management.

Even within the consent management space, there can be multiple models. For instance, some individuals may find that personal data stores,[51] where personal data is housed with an intermediary but under the control of the individual, offer a more useful mechanism than DEPA's consent managers. Others may find that the data blindness feature of consent managers does not suit their needs. They may prefer to avoid the cognitive burden and fatigue of dealing with multiple actors by using a trusted infointermediary[52] who may act as an information broker on their behalf, offering customised advice and services

49 Rohan Jahagirdar and Praneeth Bodduluri, 'Digital Economy: India's Account Aggregator System Is Plagued by Privacy and Safety Issues' (2020) 55 Economic and Political Weekly <https://www.epw.in/engage/article/digital-economy-indias-account-aggregator-system> accessed 26 October 2022; Sangh Rakshita and Shashank Mohan, 'Comments To NITI Aayog On The Draft Discussion Paper on Data Empowerment and Protection Architecture' <https://ccgnludelhi.files.wordpress.com/2020/11/ccg-nlu-comments-to-niti-aayog-on-the-draft-discussion-paper-on-the-data-empowerment-and-protection-architecture.pdf> accessed 26 October 2022; Shweta Reddy and others, 'The Centre for Internet and Society's Comments and Recommendations to the: Data Empowerment and Protection Architecture' <https://cis-india.org/depacomments1> accessed 26 October 2022; Vikas Kathuria, 'Data Empowerment and Protection Architecture: Concept and Assessment' (Observer Research Foundation 2021) <https://www.orfonline.org/research/data-empowerment-and-protection-architecture-concept-and-assessment/> accessed 26 October 2022.

50 Liz Coll (n 16).

51 Guillaume Brochot and others, 'Study on Personal Data Stores Conducted at the Cambridge University Judge Business School' (European Commission, Directorate-General of Communications Networks, Content & Technology 2015) <https://digital-strategy.ec.europa.eu/en/library/study-personal-data-stores-conducted-cambridge-university-judge-business-school> accessed 26 October 2022.

52 John Hagel III and Jeffrey F Rayport, 'The Coming Battle for Customer Information' [1997] *Harvard Business Review* <https://hbr.org/1997/01/the-coming-battle-for-customer-information> accessed 26 October 2022.

based on their data. These are still early days for privacy tech in India and the doors needs to be left open for all competing technical solutions and models.

To be fair, the existence of DEPA does not preclude the emergence of competing standards and models, but it does increase the entry barriers. In a set up where India is yet to enact a personal data law, current incentives to develop innovative technological solutions to aid compliance with its provisions are limited. These incentives would be further diminished by the fact that several government agencies have already picked a winner in terms of the technical standards for consent management. Therefore, despite DEPA having useful design elements, the manner in which it is being brought into effect can deter innovation and competition in privacy-tech in the long run. At the same time there could also be competition issues within the DEPA ecosystem — despite having a large number of players the market could evolve in a manner where a handful of large players control most of the consent transactions.[53]

4.2. Consent Management is Only a Part of the Solution

As discussed in section 4, there are a number of factors that complicate consent in the digital context. The DEPA framework speaks directly to some of those issues, mainly through its ability to maintain verifiable consent logs, with specific details like the purpose, duration, and timestamps of consent. This is a clear improvement over the opaque and sweeping ways in which consent is currently being managed. However, the consent problem is also about the fatigue that individuals face while granting numerous rounds of consent to multiple actors. It is possible that this problem would only be amplified in a system that requires granular consent each time the data is being shared or for every interaction with a new entity.[54]

As the footprint of DEPA expands to multiple sectors, and entities across sectors start interacting with one another, for instance data exchange between banks and telecom companies, the number of consent requests is likely to increase significantly. At the same time, the ability of individuals to appreciate the full consequences of the cross linkages of data may start diminishing. Therefore, the same cognitive and behavioural limitations that lead to mechanical consent authorisations in existing systems could travel into the DEPA framework. The fact that this model is being endorsed through regulatory mechanisms could also induce a placebo effect of protection and empowerment.[55]

---

53 Jain (n 46).

54 This is part of the ORGANS framework of DEPA which stands for consent that is based on Open standards, Revocable, Granular, Auditable, Notice to all parties, and Secure by design. NITI Aayog (n 2).

55 Regulatory intervention can the risk perceptions that individuals take into account while granting their consent. Heng Xu and others, 'The Role of Push-Pull Technology in Privacy Calculus: The Case of Location-Based Services' (2009) 26 Journal of Management Information Systems 135 <https://www.tandfonline.com/doi/full/10.2753/MIS0742-1222260305> accessed 26 October 2022.

Further, the terms and conditions typically contained in a privacy agreement cover a number of issues that go beyond data collection and sharing, such as data access and correction rights, retention norms, redress mechanisms, etc. Therefore, DEPA's consent practices are likely to apply in addition to, and not replace, the cumbersome terms and conditions that users have to sign up for. In fact, in her review of the service terms and privacy policies of certain Account Aggregators, Jain notes that these entities also continue to adopt policies that are 'are lengthy, full of legalese, and not very easy to understand'.[56] Moreover, even under DEPA, organisations would be the ones that decide on the scope and purpose of data and the individual's role is limited to agreeing or disagreeing with the same, albeit under a more transparent system.

4.3. Less Data is also a Form of Empowerment

Empowerment is not just about exercising agency over collected data. It can also mean having less data about ourselves available in an easily accessible digital format, which automatically means lesser exposure to data exploitation, breach and misuse. However, DEPA and related systems are designed to encourage the conversion of paper-based systems into standardised digital formats and attract more and more players into this data sharing ecosystem. For instance, the Ayushman Bharat Digital Mission, which includes consent managers as part of its design, is actively encouraging the generation of electronic health records by doctors, labs, clinics, and other establishments.[57] In addition to increased threats of data breach, the policy nudge towards greater datafication also creates vulnerabilities from increased surveillance, profiling, and potential discrimination.

While the designers of such systems may argue that both individuals and establishments have a choice of whether to participate in them, in reality, this choice is limited by many structural barriers, some of which are described below.

Given the power differential in the relationship between a doctor and a patient or a credit issuing company and a loan seeker, an individual's choice in refusing permissions for data collection or sharing in these contexts can be illusory. Even in the case of businesses, scale can eventually become a powerful driver for participation. Moreover, regulatory diktats could easily turn the system's switch from voluntary to mandatory at any point, even if the original design

56 Jain (n 46).

57 National Health Authority, 'Strategy Overview: Making India a Digital Health Nation Enabling Digital Healthcare for All' <https://www.niti.gov.in/sites/default/files/2021-09/ndhm_strategy_overview.pdf> accessed 26 October 2022.

of DEPA may not have intended that to be the case. This is evidenced by the experience from other technological systems like Aadhaar and Aarogya Setu.[58]

4.4. More Work is Needed on Data Governance

A smaller data footprint is also beneficial in light of the limited ability of individuals to detect any misuse of their personal data. The focus of DEPA has mainly been on building the technical standards for consent and sharing while relying on regulatory processes to fix other important elements of data governance. The screenshot below from Sahamati's frequently asked questions (FAQs) offers an illustration.[59] The FAQs from October 2021 contained two different responses to an identical question on what prevents a financial information user from misusing data for purposes other than those authorised by the consent artifact. The first response notes that the guidelines in this regard are yet to be framed, which is worrying since parts of the system have already come into effect. The second response points to the recommendations made by the Justice Srikrishna Committee, which do not have any binding force, and the guidelines made by sectoral regulators.[60]

Figure 2: Sahamati's FAQs on data misuse by information users (as on 21 October 2021)

- **How do you ensure that the FIU doesn't use your data for other reasons than the reasons mentioned in the Consent Artefact?**

  The FIUs will have to adhere with the Data Governance guidelines to prevent misuse of data. The guidelines are being finalised together with ecosystem players and will be shared when complete.

- **How do you ensure that the FIU doesn't use your data for other reasons than the reasons mentioned in the Consent Artefact?**

  The FIUs will have to adhere with the Data Governance guidelines to prevent misuse of data. The Srikrishna Report is the gold standard on Data Governance. Existing guidelines on security and privacy already exist for registered/regulated entities by their sectoral regulators.

58 Aarogya Setu is a contact tracing mobile application developed by the Indian government during COVID-19. 'Aarogya Setu' <https://aarogyasetu.gov.in/> accessed 8 November 2022.

59 'Sahamati | Certified Entities in the Account Aggregator Ecosystem' (n 22). The discussion that follows is based on the version of the website that was accessed on 21 October 2021. The website has since been updated to reflect only the second response shown in Figure 2.

60 The Srikrishna Committee's recommendations went through several changes under the PDP Bill, 2019 and subsequently in the recommendations of the Joint Parliamentary Committee. It remains to be seen how the next version of the bill will interact with the Srikrishna Committee's recommendations.

It is true that the RBI's directions on Account Aggregators specifically bars information users from processing data in breach of the consent artefact.[61] But the manner in which the individual is expected to detect this sort of misuse remains uncertain. Researchers who examined the technical components of DEPA have also found that there are no technical safeguards to ensure that information users do not misuse the personal data received from the Account Aggregators.[62]

4.5. Who Will Watch the Self-Regulator?

The idea of India Stack has been around since at least March, 2016,[63] which predates the first policy action on creation of consent managers (RBI's Account Aggregator directive in September, 2016).[64] Since then, active advocacy efforts have ensured that DEPA (or the consent layer of iSPIRT's India Stack framework) has been endorsed by a number of government agencies (See Table 1). Future work on development of interoperability standards and governance codes for Account Aggregators has now been handed over to a non-profit collective called Sahamati. Sahamati's functions will include designing standards, certification systems, and codes of conduct for Account Aggregators and monitoring compliance by members. It appears to be styled along the lines of the National Payments Corporation of India (NPCI), a non-profit body that owns and acts as the de facto regulator, in addition to being a dominant actor, in the UPI ecosystem.[65]

But, unlike NPCI, which derives its authorization from the Payments and Settlements Systems Act, 2007,[66] Sahamati does not have any regulatory basis. The RBI's Account Aggregator rules do not envisage the existence of such a body and as a result do not set any bounds on the powers that Sahamati can exercise viz-a-viz Account Aggregators that are regulated entities. For instance, what would be the practical consequences for a consent manager that does not

---

61 Rule 7.6.2, RBI Account Aggregator Directions (n 7).

62 Malavika Raghavan and Anubhutie Singh, 'Building Safe Consumer Data Infrastructure in India: Account Aggregators in the Financial Sector (Part–2)' (*Dvara Research Blog*, 7 January 2020) <https://www.dvara.com/research/blog/2020/01/07/building-safe-consumer-data-infrastructure-in-india-account-aggregators-in-the-financial-sector-part-2/> accessed 26 October 2022.

63 iSPIRT, 'India Stack - Towards Presence-Less, Paperless and Cashless Service Delivery' (1 March 2016) <https://www.slideshare.net/ProductNation/india-stack-towards-presenceless-paperless-and-cashless-service-delivery-an-ispirt-initiative> accessed 26 October 2022.

64 An in principle decision to move in this direction had already been taken in a 2014 meeting of the Sub Committee of the Financial Stability and Development Council but the specific details of the design and protocols that may have been discussed in that meeting are not available publicly. Reserve Bank of India, 'Press Release: 13th Meeting of the FSDC Sub Committee' <https://www.rbi.org.in/scripts/BS_PressReleaseDisplay.aspx?prid=31818> accessed 26 October 2022.

65 Advait Palepu, 'Deciphering NPCI's Dominance in Digital Payments' (*MediaNama*, 28 October 2020) <https://www.medianama.com/2020/10/223-deciphering-npcis-dominance-in-digital-payments/> accessed 26 October 2022.

66 The NPCI was set up with the support of the RBI and the Indian Banks' Association to act as an umbrella organisation for the retail payments system in India. It is authorised to act as such under Section 4 of the Payments and Settlements Systems Act, 2007. Reserve Bank of India, 'Reserve Bank of India, Booklet on Payment Systems, Chapter 3 - Institution Building - Umbrella Organisation' <https://rbi.org.in/scripts/PublicationsView.aspx?Id=20315> accessed 25 October 2022.

become a member of Sahmati? It is also not clear if Sahamati's mandate might later be extended to cover newer consent management frameworks that come up in other sectors. This would only strengthen the need for accountability, transparency and due process in the functioning of such a body.

Conclusion

Despite its many limitations, the requirement of informed consent for the collection and processing of personal data constitutes a core pillar in any data protection framework. Better management of consent, which is what DEPA seeks to achieve through its verifiable consent logs, standardized formats and data sharing APIs, can therefore be an important pathway to data empowerment. It would, however, be rash to imagine that there can be only one particular pathway to empowerment, be it through the enactment of a new law or the introduction of a new technical architecture. Data empowerment is a far-reaching goal, one that will require a multifaceted approach, focused on improving consent, building trust, and ensuring effective accountability.

Given this context, it seems problematic to see one particular model of consent management and data sharing being endorsed by the state across multiple sectors. This can deter the growth of alternate standards and models, some of which could possibly turn out to be more empowering than the proposals currently under deployment. It seems particularly dangerous to go down this path even before the evidence from DEPA's first use case, as the Account Aggregators framework in the financial services sector, comes to light.

Besides issues of state endorsement, innovation and competition in privacy-tech, this article highlights the need for humility in understanding the various factors that complicate consent in the information age. This includes several cognitive, behavioural, and structural barriers, many of which will continue to exist, and some others may be added, with the implementation of DEPA. Accordingly, more work is needed in terms of deciphering these complications while also facilitating accountability mechanisms that will apply over and above informed consent. Developing technical safeguards to enable users to track actual usage of their data and alert them of applications that go beyond the purposes authorized under the consent artifact could be a part of the solution.[67]

As the DEPA mechanism grows in scale, so will the power of the entities responsible for its design and implementation. A discussion on the accountability, transparency,

67 Jain (n 46); Malavika Raghavan and Anubhutie Singh (n 62).

and due process in the functioning of bodies such as Sahamati, the self-organised self-regulatory body of Account Aggregators, should, therefore, become an early part of the policy conversations on DEPA.

To conclude, let me invoke the need for what Sheila Jasanoff has referred to as technologies of humility.[68] Jasanoff argues that the real problems that we encounter in the real world are inherently complex and science and technology can offer only part of the solution. For instance, we don't know what data empowerment means to different individuals and groups. Will creating incentives for more data sharing enhance or diminish empowerment? How will the behavioural, technical, business, and regulatory variables in DEPA interact in practice? In the face of these ambiguities and complexities, a non-state backed, gradual and dynamic approach to data management might be preferable to the trajectory currently being pursued by DEPA.

68 Sheila Jasanoff, 'Technologies of Humility' (2007) 450 Nature 33 <https://www.nature.com/articles/450033a> accessed 25 October 2022.

Group Data Rights in Law and Policy

Arindrajit Basu and Amber Sinha[1]

Introduction

In the age of big data and analytics, data is no longer collected only about one individual or even a small group of individuals but increasingly about large and, often undefined groups.[2] Data is then transformed through multiple layers of algorithmic processing into patterns and group profiles that are applied on a macro-scale. While the individual may not be the focal point of algorithmic processing and its derivative value, fundamental rights, including the right to privacy remain vested in individuals as a fundamental tenet of any democracy.

In addition to individual data rights, over the last few years, the concept of "community data" has been proposed by Indian policy-makers both as a means of dismantling the monopoly of Big Tech companies over data, and to empower "communities" by offering them some control over data that they generate.[3] This legal innovation poses several questions ranging from how the idea of community data is compatible with the fundamental right to privacy and group privacy to how it can be operationalised to empower communities as bearers of civil, political and socio-economic rights with respect to the data they generate.[4] Further, this innovation is plagued with several critical definitional uncertainties— most pressingly, how a community is to be defined, and when personal or non-personal data can be classified as "community data."

The report on the Data Protection Bill, 2021 by the Joint Parliamentary Committee of the India Parliament includes within its scope not merely personal data, but also non-personal data.[5] The primary stated rationale for the widening of the scope and ambit of what was intended as personal data protection regulation is that distinguishing between personal data and non-personal data can be difficult with the emergence of digital technologies. The Joint Parliamentary Committee's report reads more as an expression of regulatory intent than prescriptive substantive provisions on

---

1 Amber is currently Senior Fellow for Trustworthy AI at Mozilla Foundation. He can be reached at ambersinha07@gmail.com. Arindrajit is a Non-Resident Research Fellow, Centre for Internet and Society, India. He can be reached at arindrajit@cis-india.org.

2 Linnet Taylor, Luciano Floridi, and Bart Van Der Sloot, 'Introduction: A New Perspective on Privacy' in Linnet Taylor, Luciano Floridi, and Bart Van Der Sloot (eds), *Group Privacy: new challenges of data technologies* (Dordrecht:Springer 2017).

3 'Non-Personal Data Governance Framework' <https://prsindia.org/policy/report-summaries/non-personal-data-governance-framework>.

4 Udbhav Tiwari, 'India's Ambitious Non Personal Data Report Should Put Privacy First,for Both Individuals and Communities' (*Mozilla*, 12 September 2020) <https://blog.mozilla.org/netpolicy/2020/09/12/indias-ambitious-non-personal-data-report-should-put-privacy-first-mozilla/>.

5 Report of the Joint Committee on the Personal Data Protection Bill, 2019 (16 December 2021), <http://loksabhaph.nic.in/Committee/CommitteeInformation.aspx?comm_code=73&tab=1&gt&gt>.

how non-personal data ought to be regulated and protected. While we do not agree with the reasoning for widening the ambit of the proposed law on personal data protection and extending the ambit of a privacy legislation to provisions concerning mandatory sharing of non-personal data would be inappropriate, we believe that this development provides an interesting intersection with the subject of this paper—the articulation of group rights over data, and the consequent enforcement of individual rights therein. The state of legal and regulatory theory of the governance of non-personal data is still in its early stages, and we intend for this paper to be a useful contribution towards thinking through the considerations that should underlie any such efforts. While dealing with the limited question of group rights over data, we also attempt to articulate a prescriptive hierarchy of these considerations.

For starters, the use of the term 'community' and its juxtaposition with 'data' causes unnecessary terminological confusion. If we agree that a community is a 'group of people,' then using the phrase 'group' and juxtaposing it to form 'group data rights' is broader and allows us to duck the unnecessary question of when a 'group of people' form a 'community. Academic scholarship also refers to collective rights as 'group rights'[6] and privacy scholars refers to collective privacy rights as' group privacy.'[7] Therefore, throughout this essay, we use the term 'group' unless we are directly quoting from other sources, such as the Non-Personal Data report that uses the word 'community.'

By locating group rights in theoretical constructs, Indian constitutional frameworks, and international law, we engage with the relationship between groups and the rights they hold, and the framework within which it can exert rights over data. We also engage with the relationship between groups and its composite individuals as well as external actors. Finally, we look into the pressing question of algorithmically determined groups where its composite individuals may not even be aware of their membership, and explore how groups could still serve as a unit of harm mitigation.

1. Defining a Group

After a series of unclear policy ventures into notions of data as a societal commons or an actionable interest for a community, the revised Non-Personal Data Report (hereinafter 'NPD report') issued by the Committee of Experts at the Ministry of Electronics and Technology in December 2020 attempts to bring some definitional clarity.[8] It defines a community as a "group of people that are bound by common

6 'Group Rights' [2022] Stanford Encyclopaedia of Philosophy <https://plato.stanford.edu/entries/rights-group/#:~:text=A%20group%20right%20is%20a%20right%20possessed%20by%20a%20group,people%20to%20be%20self%2Ddetermining>.

7 Linnet Taylor, Luciano Floridi, and Bart Van Der Sloot (n 2).

8 The first report had been published for public consultation in June 2020. The revised report is availableCommittee of Experts, 'Report by the Committee of Experts on Non-Personal Data Governance Framework' (The Ministry of Electronics & Information Technology 2020) <https://ourgovdotin.files.wordpress.com/2020/12/revised-report-kris-gopalakrishnan-committee-report-on-non-personal-data-governance-framework.pdf&gt>.

interests and purposes, involved in social and/or economic interactions, or other societal interests and objectives, and/or an entirely virtual community."[9] This sweeping definition is accompanied by the assertion that once personal data is anonymised or data pertains to things other than a person, such as a natural phenomenon, there is no specific data principal. The report goes on to argue that benefits from data should accrue not only to the organisations that collect and process this data but equally to "India and the community that typically produces the data that is being captured."[10] As per the report, rights over non-personal data include the right to derive economic or other value and maximise data's benefits for the community and a right to eliminate or minimise harm to that community.[11]

These assertive policy prescriptions provide plenty of opportunity for concrete deliberations that spur ahead a more equitable digital ecosystem. However, to do so, several assumptions and assertions in the report's framing need to be debunked and unpacked.

What are the groups of people that may exist then?[12] A group could be self-aware and come together through a common identity that individuals within the group ascribe to. This includes cultural or religious identities where every individual member of that group is aware that they are a member of that said group. Such groups may also possess internal structures, rules and decision-making processes such as the Lok Sabha, the faculty of universities or the Indian National Congress—what French (1984) terms a 'conglomerate collectivity.'[13] A group could also be a mere set of individuals who are brought together by factors other than a common identity that they espouse, what French terms 'aggregate collectivities.' This could include the attendees at a cricket match or those who bought furniture at a specific store on a certain date. An aggregate collectivity could also be externally imposed on individuals, such as a statistical measure of individuals below a state demarcated poverty line or a group algorithmically determined to be more likely to commit crime. In such cases, where the creation of the group itself is externally imposed, algorithmically or otherwise, the levers of control are not with the individuals that make up the collective or even the collective at large, but with an external entity that could be the state or a private actor. In a pre-algorithmic world, aggregate collectives were difficult to identify and target. Data generation, curation and algorithmic processing has made possible newer and possibly more intrusive forms of identification that drive public policy and commercial decisions alike in the modern day. As we discuss later in the essay, for instances of algorithmically created aggregate collectives, individuals in that collective such as those that are targeted with specific advertisements may not even be aware of the existence of this externally imposed group membership.

---

9 Ibid 16.

10 ibid 6.

11 Ibid.

12 'Group Rights' (n 6).

13 Peter A. French, *Collective and Corporate Responsibility* (New York: Columbia University Press 1984).

Taking cue from the NPD Report itself, groups can be defined for the purpose of guaranteeing rights that the group may exercise as a collective as well as for the purpose of preventing harms. This leads us to three core questions that we answer in this Essay:

A. When should data rights vest jointly in a group rather than separately from rights vesting in the individuals making up the group?

B. How would data rights vesting in the group play out *qua* external entities including other groups and *qua* the individuals making up the group?

C. How could the enforcement of rights and prevention of harms be ensured in the case of algorithmically determined aggregate collectives?

2. Group Data Rights and Value Aggregation in Conglomerate Collectives

Group rights can be understood through two forms of what we call 'value aggregation.'[14] The first, rarer, form occurs when it is "intrinsically valuable" for a right to vest in a group and not in the individuals making up the group. For example, the right to self-determination guaranteed by Article 1 of International Covenant on Civil and Political Rights (ICCPR) clearly vests in 'peoples' and not individuals. While theoretically, individuals may have a right to national self-determination, it is practically not possible for this right to be enforced. The *sequitur* here is, as Raz and Margalita put it, "that the moral importance of the group's interest depends on its value to individuals,"[15] even if the right vests in the group and not in the individual.

A second more common form of 'value aggregation' occurs when groups are "bearers of value" for the enforcement and actualization of individual rights or for the enjoyment of 'participatory goods.' In such cases, the group does not have any rights independently of its composite individuals but instead is an intermediary for the enforcement of individual rights. *Value aggregation* in the second instance does not accord rights to a group *per se*. Examples include the rights granted to religious or linguistic minorities, a right to enjoy with other members of the group the enjoyment of their culture, use of their language and practising of their own religion. Religion, culture, and language are all key examples of participatory goods, whose value lies significantly in its shared enjoyment. The right to speak one's language would be far less valuable if others cannot share, understand and exchange this experience. However, this reality does not alter the nature of the right, which clearly vests in

14 Miodrag A. Jovanovic, *Collective Rights: A Legal Theory* (Cambridge University Press 2012). Jovanovic uses the phrase 'value collectivism' to also conceptualize cases where group rights may subsume or supersede individual rights, which our framing of 'value aggregation' does not endorse.

15 Avishai Margalit and Joseph Raz, 'National Self Determination' (1990) 87 The Journal of Philosophy 87.

individuals and not in groups, even though the beneficial interest for individuals lies in shared enjoyment.

Article 27 of the International Covenant on Civil and Political Rights indicates that it is the 'persons belonging to such minorities "who are the bearers of rights, but the other members of the group are integral to the value underpinning that right. General Comment no. 23 on minority protection stipulates that "[a]lthough the rights protected under Art.27 are individual rights, they depend in turn on the ability of the minority group to maintain its culture, language or shared religion. Accordingly, positive measures by the state may also be necessary to protect the identity of a minority."[16] Section 31 of the South African Constitution is framed in much the same way. The text of the Indian Constitution takes a different approach in Articles 25 and 26 where it makes religious groups the express bearers of rights in Articles 25 and 26, although as we show in the next section, it envisages the second form of value aggregation and not the first.[17]

When it comes to the vesting of data rights, the second form of value aggregation is far more appropriate as we see several practical instances where groups may be useful intermediaries for the enforcement of rights but the right itself vests in individuals and not the group.

As a modified application of Raz's work on collective rights, we propose the following criteria for groups to be seen as intermediaries for the enforcement of individual rights or the enjoyment of public goods[18]:

A. Individual members must perceive themselves to be normatively bound to each other or to have collective interests that they share with the group.

B. The group is working in adherence with, and for the advancement of, this shared normative understanding which could include decision-making processes, membership rules, adjudication mechanisms and other factors integral to the enforcement of rights or enjoyment of participatory goods.

C. Individual members of the group pooling their collective interests believe that doing so will lead to value aggregation either in the enforcement of rights or in the enjoyment of participatory goods.

As McDonald argues, the legal recognition of a group is not necessary for a group to exist, but the law should endeavour to recognize groups that exist through

16 UN Human Rights Committee (HRC), *CCPR General Comment No. 23: Article 27 (Rights of Minorities)*, 8 April 1994, CCPR/C/21/Rev.1/Add.5, para 6.2.

17 Gautam Bhatia, *The Transformative Constitution: A Radical Biography in Nine Acts* (Harper Collins Publishers India 2020) 163.

18 Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon Press).

the social fact of shared normative understandings.[19] Our definition adds several criteria to that proposed in the NPD report-the requirement of shared normative understandings, the requirement that the group is working to advance and uphold these understandings, and the requirement that individual members must see value in pooling their interests with that of the group.

A potential example of a group that benefits from value aggregated data rights is that of a data co-operative. A co-operative, defined as "autonomous association of persons united voluntarily to meet their common economic, social, and cultural needs and aspirations through a jointly-owned and democratically controlled enterprise"[20] forms when individuals feel that there are collective interests that can be more effectively realised if resources are pooled. Data co-operatives are co-operative organisations that are formed for the stewardship of data for the benefit of their members who themselves are individuals (or data subjects, such as SalusCoop, a health data co-operative whose members can review medical research proposals and consent to sharing data for specific medical research projects[21] However, before going further with this framework, we must understand the potential conflict of rights that it could lead to-critically between individuals making up the group and the group itself.

3. Group Data Rights qua External Actors and Individuals in Conglomerate Collectives

Group data rights may be an effective means of ensuring individual rights through several avenues. Taking the example of health data sharing, Sridharan argues that user-centric models of stewardship through cooperatives can empower individuals and communities to exercise decisional autonomy.[22] Sridharan uses the example of SalusCoop,[23] a health data cooperative whose members can review specific medical research proposals and consent to sharing data for specific research projects. Decisions are made through democratic processes with members being accorded one vote each. Another example used is Citizen[24]- a private collaborative platform that enables both individual patients and larger patient advocacy groups to share health information for research by pharmaceutical companies. With Citizen, individual consent is essential with each specific research project. Group data rights could lead to value aggregation in terms of the enforcement of rights by augmenting transparency,

19 Michael McDonald, 'Should Communities Have Rights? Reflection on Liberal Individualism' (1991) IV Canadian Journal of Law and Jurisprudence 218.

20 Home 'International Cooperative Alliance' <https://www.ica.coop/en>.

21 Ada Lovelace Institute, 'Exploring Legal Mechanisms for Data Stewardship' (Ada Lovelace Institute) <https://www.adalovelaceinstitute.org/feature/data-cooperatives/#fnref-2>.

22 Soujanya Sridharan, 'Health Data Governance: Empowering Communities to Effectively Manage Their Data' (*Aapti Institute*, 24 May 2021).

23 'Salus Coop' <https://www.saluscoop.org/acerca>.

24 'CitizenMe' <https://www.citizenme.com/>.

vouching for the conduct of regular audits, and increasing the collective bargaining power of users against external actors-private data processors and the state.

At the same time, we are left with the pressing question—what if an individual wants to override decisions made by the co-operative with respect to the sharing of their data? Let us consider an instance, where a co-operative democratically decides to go ahead with sharing the (anonymised) data of that co-operative with an external private processor but an individual in the minority remains uncomfortable with that decision. Would the individual right to privacy override the individual's participation in that data co-operative? We argue that the answer should lie in the affirmative, given the personal and inalienable nature of the right to (informational) privacy, the final reins of control, including a right to opt-out should vest with the individual at all stages of the data cycle.

Our argument is rooted in the conceptualization of rights both in international and constitutional law. As discussed above, in international law, individuals, not groups are bearers of rights. The Indian Constitution, through Article 25(b) makes groups the direct bearers of rights. However, as Bhatia rightly opines[25]

[Article 26(b)] does not clarify whether groups are granted rights for the instrumental reason that individuals can only achieve self-determination and fulfilment within the context of choice provided by communities or whether the Constitution treats groups, along with individuals as constitutive units worthy of equal concern and respect.

When answering this question, Bhatia cites Ambedkar who specifically argued that Indian constitutionalism and the architecture of individual rights conflicted not only with the state but also with hierarchical social relations fermented by 'self-regulating communities.'[26] Therefore, Bhatia concludes that the constitutional vision sees groups as bearers of value but does not grant them constitutive value which would override individual claims. Thus, the cultural survival of groups is an important derivative right but neither an end in itself nor a substitute for individual rights.

A similar paradigm exists on the import of the right to privacy. This was discussed at some length by Justice Chandrachud in *KS Puttaswamy (I) v Union of India* through the prism of decisional autonomy. He cited the Delhi High Court judgement that had held in *Naz Foundation* that the sphere of privacy allows persons to develop human relations and exercise autonomy without outside interference from both the outside community and the state. Justice Chandrachud further explicitly proclaimed that the "individual is the focal point of the Constitution because it is in the realisation of

25 Gautam Bhatia, 'Freedom from Community: Individual Rights, Group Life, State Authority and Religious Freedom under the Indian Constitution' (2016) 5 Global Constitutionalism 351.

26 Gautam Bhatia (n 17) xxvii.

individual rights that the collective well-being of the community is determined."[27]

This means that a critical fourth criteria must be added when groups are considered intermediaries for individuals - that of free and informed consent and a right to opt-out for all individuals choosing to pool their resources as part of a group.

The NPD report itself adopts an interesting approach in this regard. Given the risks of anonymised personal data being re-identified, there is a valid debate on the extent to which anonymised data sets receive protection under data protection laws. The General Data Protection Regulation (GDPR) through Recital 26 adopts a risk-based approach to determine whether data is personal or not, an approach used by the UK Information Commissioner's Office (ICO) as well.[28] This approach entails a risk assessment. If it shows that identification is 'reasonably likely' to occur, anonymised data receives GDPR protection fully. However, the Article 29 Working Party of the European Union suggests a higher threshold - arguing that anonymised personal data can only qualify as non-personal data when the anonymisation is irreversible - technically a hugely challenging threshold for any data processor to muster.[29] The NPD report, on the other hand, opts for a more nuanced approach than the contrasting European approaches. It recognizes the challenges of irreversibly anonymising datasets and rather than setting an impossible threshold for anonymisation, it roots the solution to this problem in informed consent, including requirements of disclosure, notice that the personal data will be anonymised and an opt-out mechanism for the data principal.[30]

These principles should also apply when adjudicating conflicts between individuals and groups, such as data co-operatives that individuals may have opted into. The key problem stemming from collectivising value is the possibility of hierarchisation within the group itself-where an individual or a subset of individuals start taking decisions that harm other individual members of the group. While individuals may choose to pool their data in and safeguard rights or derive other beneficial interests through the group, the individual nature of the right to (informational) privacy can never be compromised.

4. Algorithmically Determined Aggregate Collectives

The previous sections dealt with conglomerate collectives - self-aware groups where individuals consent to operationalizing their rights through a group such as a data

---

27 *Justice K.S.Puttaswamy (Retd). v. Union of India And Ors.* (2017) 10 SCC 1, para 96.

28 Michele Finck and Frank Pallas, 'They Who Must Not Be Identified—Distinguishing Personal from Non-Personal Data under the GDPR' 10 International Data Privacy Law 11.

29 Article 29 Working Party, 'Opinion 05/2014 on Anonymisation Techniques' 11–12, 23–25.

30 Amber Sinha and Arindrajit Basu, 'Community Data and Decisional Autonomy: Dissecting an Indian Legal Innovation for Emerging Economies' (*Medium*) <https://medium.com/digital-asia-ii/community-data-and-decisional-autonomy-dissecting-an-indian-legal-innovation-for-emerging-409157fd7788>.

co-operative. In this section, we deal with cases where groups are externally imposed on individuals through data analytics and algorithms. Algorithmic classification has two possible implications for group formation and group rights. First, profiling could be used to draw inferences about conglomerate collectives. This includes inferences about groups formed on the basis of racial, religious or other social identities.[31] The rights of these groups against algorithmic discrimination are set out in liberal democratic constitutions, including the Equality Code in the Indian Constitution.[32] As discussed above, each individual within these groups have an individual right against algorithmic discrimination but the group serves as an intermediary for the enforcement of these rights.

Second, analytical tools create aggregate collectives without consent or even awareness of the individuals.[33] Studies show that users either feel resigned to being tracked and are unaware of the extent of commercial surveillance, including ad profiling that they are subjected to.[34] Google's ad preferences for example creates customised advertisements for all digital profiles based on factors information in one's Google account, including age range and gender, location, search history, activities while the user was signed into Google, website history, mobile applications and activities on any other device they may own.[35] These troves of data are curated and algorithmically processed to create digital profiles for each user that influences the targeted advertisement they receive.[36] This includes details like parental status, household income, job industry, areas of interest such as 'politics', 'pop music' or 'American Football.'[37] This is just the tip of a problematic iceberg. With Google, it is at least possible to track one's digital profile and opt out of receiving targeted advertisements. In most other cases, data consensually shared with one platform is surreptitiously syphoned off to third-party data brokers who subsequently sell it to other parties.[38]

As Tanya Kant eloquently puts it, "it is not just individualistic selves who are

31 Virgina Eubanks, *Automating Inequality: How High Tech Tools Profile, Police and Punish the Poor* (St Martin's Press 2018).

32 The Constitution of India, 1950, articles 14-18.

33 Caroline Bassett, 'Identity Theft' in Caroline A. Jones (ed), *Sensorium: Embodied Experience, Technology and Contemporary Art* (Cambridge, MA: MIT Press, 2006).

34 'Adults' Media Use & Attitudes' (OfCom 2020).

35 'My Ad Centre' (*Google*) <https://myadcenter.google.com/?sasb=true>.

36 Tanya Kant, 'Identity, Advertising, and Algorithmic Targeting: Or How (Not) to Target Your "Ideal User"' [2021] MIT Case Studies in Social and Ethical Responsibilities of Computing <https://mit-serc.pubpub.org/pub/identity-advertising-and-algorithmic-targeting/release/2>.

37 Based on a selection of one of the author's listed ad preferences-several of these profiles are inaccurate.

38 As told by Mishi Choudhary to The Economic Times, "When you sign up for free discounts, fill out questionnaires, or your clickstream in general, you are giving up all the data voluntarily and agreeing to privacy policies that allow you to do so." Aritra Sarkhel and Neha Alawadhi, 'How Data Brokers Are Selling All Your Personal Info for Less than a Rupee to Whoever Wants It' *Economic Times* (28 February 2017) <https://economictimes.indiatimes.com/tech/internet/how-data-brokers-are-selling-all-your-personal-info-for-less-than-a-rupee-to-whoever-wants-it/articleshow/57382192.cms> ; See also Ronald J. Deibert, *Reset: Reclaiming the Internet for Civil Society* (House of Anansi Press 2020) 104–147; See also for more on advanced data brokerage markets and their harms in the US and North America Justin Sherman, 'Data Brokers and Sensitive Data on U.S. Individuals' (Sanford School of Public Policy 2021) <https://sites.sanford.duke.edu/techpolicy/wp-content/uploads/sites/17/2021/08/Data-Brokers-and-Sensitive-Data-on-US-Individuals-Sherman-2021.pdf>.

managed, reduced and verified through data, collective audiences are also reduced and reshaped through algorithmic sorting and auditing techniques."[39] Therefore, even though such algorithmically determined aggregate collectives may not be bearers of rights or even allow for the explicit aggregation of value but nonetheless need to be discussed for the mitigation of harms that may arise to individuals making up that collective.

Socrates believed that the unexamined life is not worth living and put a high premium on self-awareness and self-knowledge.[40] His emphasis on a life of inquiry and a life of reason, must be read with his argument that all people only desire the good. The central assumption in this argument is that no one desires to harm themselves. Since desiring what is bad is wishing to secure something harmful for oneself, and securing what is harmful for oneself is harming oneself, no one ever desires what is bad.[41] It follows that individuals only harm themselves by acting for something that they think is good, but is really bad. Only when acting from a false claim of knowledge (of what is good) can people harm themselves. While Socrates may have regarded virtue as a matter of knowledge and vice as a matter of ignorance, in the definitions of human autonomy, this has been looked at in a much more limited context.[42] Still, it is worth noting that the idea of the informed self being tied to autonomy rises from ancient Greek philosophy. We will note the contours of meaningful autonomy for our purposes below.

The three stages of human rights that James Griffin refers to in his books, 'On Human Rights' are important in this context.[43] The first stage comprises our ability to consider our lives as a whole and reflect upon what makes our life worthwhile and to make decisions about the sort of life we want to lead. In order for us to exercise this capacity, we require autonomy. The second stage comprises those various elements that make possible the pursuit of this conception of the good life: the skills, resources and support we need to enable us to exercise autonomy are welfare provisions above some minimal level. The third stage comprises the freedom to employ our welfare provision in the exercise of autonomy, unhindered by interference from others: namely, liberty. In particular, we are interested in the second stage in which he refers to the skills, resources and support we need to enable us to exercise our autonomy. It is in this stage that Griffin seems to identify a right to minimum information. He suggests that information is a prerequisite for an individual to make real choices and be autonomous. A standard of acceptable autonomy must include the ability

---

39 Tanya Kant (n 36).

40 Thomas C Brickhouse and Nicholas D Smith, 'Plato's Socrates' [1994] New York: Oxford University Press 201.

41 ibid 92.

42 For reference, please see Christman J, 'Autonomy and the Challenges to Liberalism: New Essays' [2005] Cambridge University Press;Gerald Dworkin, 'The Theory and Practice of Autonomy' [1988] New York: Cambridge University Press; Fabian Freyenhagen, 'Autonomy's Substance' (2017) 34 Journal of Applied Philosophy 114, 114–129.

43 James Griffin, *On Human Rights* (Oxford University Press 2008).

of individuals to identify goals and ends.[44] It is in furtherance of this idea that we argue that the ability to make autonomous decisions hinges upon having access to sufficient information and further, on being able to act based on that information.

It would be tempting to look at Griffin's argument that autonomy is relative and certain kinds of losses of autonomy do not lead to an abnegation of human agency. Griffin himself argued one could make a rational choice to rely on others in many circumstances without compromising on their agency, in any real sense.[45] Similarly, complex algorithms could be relied upon to help an individual make decisions, say, about their investments. One could argue that technology has always existed which laymen have barely understood. However, while we may not have had the wherewithal to engage with minute aspects of technology, we have had the rough knowledge required to use it. Donald Norman, cognitive scientist and usability engineer, referred to this understanding as the conceptual model, and defines it as "an explanation, usually highly simplified, of how something works. It doesn't have to be complete or even accurate as long as it is useful."[46]

When we consider algorithmically determined aggregate collectives, it raises fundamental questions about the informed selves. In most cases, individuals who are classified as part of groups by algorithmic systems are not aware of such classification unless it correlates directly to groups that they perceive themselves as a part of. In certain cases, classification and sorting decisions may correlate with established and understood identities such as race, religion, gender, demographic breakups dependent on age, location and socio-economic status. Further, these algorithmically determined collectives may be used by external actors, including both state and the private actors to make key decisions such as location-based policing[47], credit rating[48], and the distribution of welfare benefits[49] in a manner that is discriminatory without the target ever getting to know they are being discriminated against.

Applying Griffin's analysis, the first policy problem to further individual agency is to address the transparency problem. As members of algorithmically determined aggregate collectives, the first challenge that individuals face are unaware of classification decisions about them, and ways in which it interferes with their agency. The most obvious risk of harm that arises from such classification is that of discrimination. All classification tasks face the challenge of achieving utility in

---

44 Robert Johnson and Adam Cureton, 'Kant's Moral Philosophy' [2019] The Stanford Encyclopedia of Philosophy <https://plato.stanford.edu/archives/spr2019/entries/kant-moral/>.

45 James Griffin (n 43) 152.

46 Donald Norman, *The Design of Everyday Things* (New York: Basic Books 2013).

47 Vidushi Marda and Shivangi Narayan, 'Data in New Delhi's Predictive Policing System' [2020] FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency <https://dl.acm.org/doi/abs/10.1145/3351095.3372865>.

48 Talia B. Gillis, 'False Dreams of Algorithmic Fairness: The Case of Credit Pricing' <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3571266>.

49 Tapani Rinta-Kahila and others, 'Algorithmic Decision-Making and System Destructiveness: A Case of Automatic Debt Recovery' 31 European Journal of Information Systems 1.

classification for some purpose, while at the same time preventing discrimination against protected population subgroups. For instance, being aware of correlations between classification decisions and protected population subgroups will help individuals understand how their rights to equality and against discrimination may be impacted. A clear example of the application of this right could be the 'but-for' test evolved through English case law.[50] Under this test, the intention, motive, reason or purpose behind an allegedly discriminatory act becomes irrelevant and the only determining factor is whether the same treatment would have been meted out but for an injured party's protected characteristic. The strength of the correspondence between the ground of distinction and protected characteristic is the key question, and it is a useful lens to study the use of proxy data for decision making. In the case of algorithmic decisions which run the risk of having indirect discrimination, this could be a guiding principle.

Even in cases where clear legal tests are harder to apply, the evolving literature on classification norms[51] can aid individuals in arriving at a conceptual model of how their rights are being impacted by algorithmic decision making. In this respect, evolving frameworks for fairness in classification are useful. Dwork et al try to address this problem from the point of view of a task specific similarity metric describing the extent to which pairs of individuals should be regarded as similar for the classification task.[52] Unlike the 'but-for' tests which would be used to retrospectively evaluate individual classification decisions, norms such as fairness through awareness attempt to integrate practices which prevent abuse of sensitive personal data in the classification step. This would mean norm-setting or regulation at the level of a metric which could have multiple classification schemes within it. The two kinds of responses mentioned here are distinct — retrospective evaluation using a principle such as 'but-for' test is suitable for regulators and adjudicators analysing the impact of algorithms; on the other hand, fairness through awareness is intended for integration in the product development lifecycle. In both the cases, the clear route towards accountability and the mitigation of harms is through algorithmic transparency in the form of clear guidance on how classification and sorting systems categorise individuals as part of groups. This need for transparency is especially critical where the aggregate collectives formed through algorithmic profiling have no correlation with the conglomerate collectives or protected sub-groups protected under constitutional law. Even in those cases, awareness about algorithmic decisions turning on the basis of perceived membership of such groups is useful for individuals to understand how decisions are made about them. This would be an essential first step towards understanding what rights, entitlements

50 *James v Eastleigh Borough Council* [1990] 2 AC 751.

51 Ninareh Mehrab and others, 'A Survey on Bias and Fairness in Machine Learning' arXiv <https://arxiv.org/pdf/1908.09635.pdf>.

52 Cynthia Dwork and others, 'Fairness Through Awareness' [2011] Arxiv <https://arxiv.org/abs/1104.3913>.

and protections may be applicable to them. After this first step of developing shared understandings or identifying collective interests, individuals may choose to pool in these interests with other similarly placed individuals to collectively bargain for their interests.

Conclusion

While the NPD report does not clarify this, it is possible to look at the proposed legal innovation of 'community data' as a response to three problems currently faced by several data protection regimes. The first relates to the difficulty in implementing the 'notice and consent' paradigm.[53] Due to structural barriers to read and understand privacy notices, failures to anticipate or comprehend the consequences of consent, and failures to opt-out, informed consent remains a broken idea which poses unrealistic expectations on individuals.

The second relates to the exploitative nature of data driven business models which privilege business profits over user agency and networks effect advantages of Big Tech over small-scale operations facing barriers to entry.[54] Given the way these business models are interwoven with even the most mundane everyday activities, the ubiquity of data collection points as well as the compulsory provision of data as a prerequisite for the access and use of many key online services, is making opting-out of data collection not only impractical but in some cases impossible. Data protection laws attempts to address this problem at the level of each specific collection of personal data, by introducing regulatory limitations when data is collected, without paying much attention to the systemic and exploitative nature of this business.

The third relates to a fundamental flaw with how individual consent is constructed. As individuals, in the most ideal of situations, we are in a position to make informed choices about ourselves. However, we are increasingly being confronted by a world where critical decisions about and for us made by data driven systems are dependent as much on choices made not just by us but others who belong to the aggregate and conglomerate collectives that we are a part of.[55] There are fundamental limitations in how effectively individual rights can address this problem.

The idea of group data rights is an appropriate regulatory response to these problems. The NPD report, however, focusses its entire energy on how mandatory

---

53 Claire Park, 'How "Notice and Consent" Fails to Protect Our Privacy' (*New America*, 23 March 2020) <http://newamerica.org/oti/blog/how-notice-and-consent-fails-to-protect-our-privacy/> accessed 4 November 2022.

54 Parminder Jeet Singh, 'Data and Digital Intelligence Commons (Making a Case for Their Community Ownership)' <https://itforchange.net/sites/default/files/1673/Data-commons.pdf>.

55 See generally Pedro Domingos, *The Master Algorithm* (Penguin Books 2017); Linnet Taylor, Luciano Floridi, and Bart Van Der Sloot (n 2); Linnet Taylor, Luciano Floridi, and Bart Van Der Sloot, *Group Privacy: New Challenges of Data Technologies* (Linnet Taylor, Luciano Floridi, and Bart Van Der Sloot eds, Dordrecht: Springer 2017).

data sharing regimes can be created, without clearly analysing the ethical basis for such mandates. A more useful account of group data rights needs to focus on the following questions.

A. *How may a self-aware group organise itself to assert group rights over data it generates?*

   Through individuals opting into self-aware groups (conglomerate collectives) such as data co-operatives, the collective can be used as a means of ensuring the secure and transparent sharing of pooled data with external actors, thus reducing harm to individuals who have their data exploited due to the present extractive nature of the digital economy.

B. *How may conflicts between such groups and individuals who are part of the group may be resolved?*

   We have argued in this paper that in the case of direct conflicts between the exercise of individual rights and group rights, individual rights should always prevail. This may be done by prioritising autonomous rights of individuals over institutional rights, such as a women's right to bodily integrity overriding state or religious interests in institutions of marriage; and individual choice over group choices in case of self-aware group. When it comes to group rights over data in the case of conglomerate collectives such as data co-operatives, this must include a right to opt-out of any instance of data sharing, even if the rest of the co-operative feels otherwise. We believe that this principle should underlie any legal regulation or self-regulatory frameworks created to enable exercise of group rights over data.

C. *What mandates are necessary to create for data processors who engage in data driven decision-making about individuals, such that individuals are able to recognise that aggregate or conglomerate collectives that they are seen as part of?*

   We argue that for aggregate or conglomerate collectives, there is a need to articulate meaningful transparency rights which enable individuals and groups to recognise their algorithmic classification in groups which leads to decisions made for and about them.

D. *What duties to assess, minimise and prevent harms to individuals as part of groups must accompany classification and sorting decisions made by data processors?*

   There is a need to article positive duties for data processors and corresponding rights for individual and conglomerate collectives. These duties must include positive obligations to prevent exclusionary and discriminatory harms to

individuals, by virtue of their membership in a group. More significantly, we recognise the need for duties in case of algorithmically determined aggregate collectives which must include an anti-discriminatory duty to ensure that prevention of indirect discrimination arising from use of proxy data for protected characteristics.

The individual right to privacy faces structural problems, both at the level of its relation with other rights, in this case group rights, and at the level of its implementation. The idea of group data rights can represent an approach which can help address some of these structural problems. By centering individual and group agency in this discourse, over business innovation and greed for indiscriminate access to data, we can help arrive at the right solutions.

Unpacking Community Data: Agency, Rights and Regulation

Kritika Bhardwaj and Siddharth Peter de Souza[1]

Introduction

Rapid advancements in data analytic techniques have spurred an important conversation regarding inadequacies of the extant data protection and governance frameworks in recent years. One such shortcoming, as highlighted by several scholars, is the failure of data protection (and more generally, governance) frameworks to recognise rights - and secure interests - of groups or communities which are impacted by processing of large scale data.[2] This gap has arisen owing to a growing consensus that big data analytics is no longer limited to profiling individuals, but is increasingly being used by businesses and states to predict behaviours of groups, which then form the basis for commercial or policy decisions affecting that group.[3] For example, location information collected from individual smartphone users, when aggregated, may be useful in deciding road or traffic policy in a particular city. These policy decisions are in-turn likely to affect all road users differently, based on whether they are pedestrians, or use private or public transport. Due to the ways in which big data processes and groups individual preferences into collective categories, there is a need to respond to the impacts and potential harms at not just an individual, but at a collective level as well.

In India, different policy documents have recently attempted to identify and articulate a need for recognising group or community rights in different data governance contexts. In 2017, a Committee of Experts came to be constituted under the Chairmanship of Justice (Retd.) B.N. Srikrishna to examine issues relating to data protection in India and to propose a draft Bill addressing them.[4] While the draft Bill proposed by the Committee did not recognise any community rights

---

1 Kritka is an independent legal practitioner. She can be reached at kritika@kbhardwaj.in. Siddharth is a Postdoctoral Researcher at the Global Data Justice Project and Tilburg Institute for Law, Technology, and Society at Tilburg University. He can be reached at S.P.deSouza@tilburguniversity.edu.

2 Linnet Taylor, Luciano Floridi, and Bart Van Der Sloot, *Group Privacy: New Challenges of Data Technologies* (Linnet Taylor, Luciano Floridi, and Bart Van Der Sloot eds, Dordrecht: Springer 2017); Joshua A.T. Fairfield and Christoph Engel, 'Privacy as a Public Good' (2015) 65 Duke Law Journal 385 <http://scholarship.law.duke.edu/dlj/vol65/iss3/1>; Salome Viljoen, 'Democratic Data: A Relational Theory For Data Governance' [2020] SSRN Electronic Journal <https://www.ssrn.com/abstract=3727562> accessed 16 November 2022.

3 Linnet Taylor, Luciano Floridi, and Bart Van Der Sloot (n 1).

4 Ministry of Electronics & Information Technology, 'Office Memorandum - Constitution of a Committee of Experts to Deliberate on a Data Protection Framework for India [No.3(6)/2017-CLES]' <https://www.meity.gov.in/writereaddata/files/MeitY_constitution_Expert_Committee_31.07.2017.pdf> accessed 15 November 2022.

over data[5], in its Report (Srikrishna Report), accompanying the draft Bill, the Committee observed that a framework for the collective protection of privacy may be required as an extension of the proposed data protection framework. It recognised the relationality aspects of data, where individual actions can influence the experiences of those around them.[6] Soon after, the 2019 draft E-commerce policy noted that subject to privacy rights being secured, a framework was required for sharing of community data in larger public interest. However, it left the idea of public interest open and evolving without specifying it.[7] More generally, India's Economic Survey 2018-19 defined data as a public good and set out its objective of harnessing large datasets for social welfare-oriented decision making. It identified that agency around the governance of data should be by the people and for people, since data is generated by them.[8] In each of these instances, despite acknowledging the importance of community data, none of the policy documents identified above provided any guidance on the constitution of a community, or a definition for community data and the governance of such data.

The most comprehensive articulation of community rights was seen in the 2020 Report of the Committee of Experts on Non-Personal Data Governance Framework, chaired by Kris Gopalakrishnan.[9] Based on feedback received from public consultations, this Committee subsequently released a revised Report (NPD Report).[10] As with the Srikrishna Report, the NPD Report acknowledged that processing of data may result in 'collective harm' to a group or community.[11] The NPD Report further observed that a framework for regulating non-personal data must allow for sharing of community data for social, public, and economic value creation.

In view of community or groups increasingly becoming the situs for data analytics and profiling, this essay argues that any regulatory framework for data governance must factor in the rights/ interests of a group or community. However, a fuller review of the existing policy proposals, in the subsequent sections of this essay, demonstrates that there is considerable policy confusion in identifying and defining a community, what community data entails, and the underlying basis for regulating and governing

---

5 While the draft Bill did not recognise community rights over data, it clarified that the provisions of the Bill shall not affect the powers of the Central Government to frame appropriate policies for data other than personal data.

6 Committee of Experts under the Chairmanship of Justice B. N. Srikrishna, 'A Free and Fair Digital Economy' (Ministry of Electronics and Information Technology) <https://www.meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf>.

7 Ministry of Commerce and Industry, 'Draft National E-Commerce Policy: India's Data for India's Development' <https://dpiit.gov.in/sites/default/files/DraftNational_e-commerce_Policy_23February2019.pdf> accessed 13 November 2022.

8 Ministry of Finance, 'Economic Survey 2018-19 - Data "Of the People, By the People, For the People"' Volume I <https://www.indiabudget.gov.in/budget2019-20/economicsurvey/doc/vol1chapter/echap04_vol1.pdf> accessed 15 November 2022.

9 Committee of Experts, 'Report by the Committee of Experts on Non-Personal Data Governance Framework' (The Ministry of Electronics & Information Technology 2020) <https://ourgovdotin.files.wordpress.com/2020/12/revised-report-kris-gopalakrishnan-committee-report-on-non-personal-data-governance-framework.pdf> accessed 4 November 2022.

10 Committee of Experts, 'Report by the Committee of Experts on Non-Personal Data Governance Framework Revised' (The Ministry of Electronics & Information Technology 2020) <https://ourgovdotin.files.wordpress.com/2020/12/revised-report-kris-gopalakrishnan-committee-report-on-non-personal-data-governance-framework.pdf> accessed 11 November 2022.

11 ibid 8.

community data. This essay therefore critically examines, in Section 1, the existing policy articulations and offers an analysis that identifies as well as addresses gaps in the aforesaid existing discourse. Section 2 of the essay sets out how some of the policy documents referred to above define a community as a unit for locating and enforcing rights. It examines definitional challenges with defining a community as a holder of rights and proposes a taxonomy of definitions based on identity and interests. In Section 3, we discuss the interconnections between community and community data by examining questions of agency and representation. The essay then progresses to look at questions of community data governance, as a framework that places emphasis on an understanding of collective harms in a regulatory context. Based on the above, we conclude the essay by arguing for a broader definition of a community and argue that any framework for the regulation of community data must secure certain basic minimum rights and outcomes.

1. Outlining the Contours of a Community

1.1. On Matters of Identity and Interest

The NPD Report defines a community as 'any group of people' bound by 'common interests and purposes' and involved in 'social and / or economic interactions.[12] It goes on to state that this could be a "*geographic community, a community by life, livelihood, economic interactions or other social interests and objectives and / or an entirely virtual community*." More importantly, the NPD Report's understanding of a community remains largely focussed on conferring it with an ability to extract economic value out of data pertaining to it.[13] The NPD Report therefore primarily ascribes value to only that data which a community may be able to commercially exploit for material wealth and well-being.[14] As elaborated upon below, this may significantly limit the scope of how communities are envisaged and recognised under any potential framework for community data as this articulation limits the formation of the community to one that has an economic rationale.

At first blush, this definition may appear to be too broad and vague, and therefore incapable of being defined in precise legal terms. However, any attempt at regulating community data must be preceded with identification of communities that have come to be formed because of ubiquitous data processing, and therefore at risk of being profiled or discriminated against. While the NPD Report expressly recognises 'virtual communities', the scope of such a community is unclear. It is important that virtual communities are

12 ibid 16.

13 ibid 58.

14 ibid.

not limited to digitised data of an already well-defined group of individuals (for example, caste data collected by the Government) but also include groups which may come to be formed only because of how captured data has been processed or applied. Such processing may inevitably end up creating newer communities which may be interested or directly impacted by it (Data Communities or a Data Community).

For instance, residents of a locality who by virtue of shopping on a particular online platform may not be aware that the online platform profiles them and offers services at a premium based on their residential status, thereby making them a separate (virtual) Data Community.

A Data Community is undeniably significantly harder to define. For one, individuals may not even be aware that their personal data, when aggregated with others', has resulted in the formation of a group based on certain behaviour or patterns.[15]

Besides the NPD Report, none of the other policy documents referred to above attempt to define a community. The Srikrishna Report only observes that an 'identifiable community' which has contributed to the body of community data, must have the right to collective protection of privacy.[16]

The idea that a group is entitled to certain special rights, or benefits owing to their distinct common identity or common interest is not new under Indian law. Articles 25 and 26 of the Constitution of India, 1950 (the Constitution) guarantee persons the right to freely profess and practice their religion; and for every religious denomination to establish and maintain institutions for religious or charitable purposes, and to manage its own affairs in matters of religion. The genesis of these group rights, therefore, stems from a recognition of the fact that persons with a common identity (e.g., the same religion) may have a common interest (e.g., to maintain religious institutions or manage religious affairs). Similarly, Article 29 recognises citizens' right to conserve their distinct language, script or culture and Article 30 confers religious and linguistic minorities with the right to establish educational institutions of their choice. The idea of a community based on religion, language or minority status is well-recognised under law. It is also easier to define, in view of the largely immutable nature of such characteristics, and an individual's membership being either inherent or express. All members of the community are also likely to be *aware* that they belong to it. In contrast, Data Communities are more fluid, in the sense that newer uses of the same data from the same individuals may create

15 Linnet Taylor, Luciano Floridi, and Bart Van Der Sloot (n 1).

16 Committee of Experts under the Chairmanship of Justice B. N. Srikrishna (n 5).

newer categories of communities within the existing Data Community. To take the example referred to in the introduction of this essay, traffic insights gained from aggregated location information from smartphones may create different Data Communities depending on whether the information is used as the basis to broaden existing roads or convert recreational areas into new roads. While the former may create an aggrieved Data Community consisting of pedestrians and cyclists, the latter may create an entirely different Data Community of residents and children.

In thinking about a community, we believe that it is important to have a criterion to build a taxonomy of different types of communities. To do this, it is important to include aspects of identity as well as interest in conceiving of a community. This distinction is important because it enables one to acknowledge that not all communities emerge with structured purpose or with a clear sanction to achieve functions. For instance, a Resident Welfare Association is created to meet the needs of a particular neighbourhood and emerges with a clear identity, a set of rules, procedures, and elected office bearers. The same neighbourhood may also have a group who are similarly interested in resident welfare particularly that of the elderly but do so by organising evening walks. The group has no fixed members, and no fixed commitments. Each configuration consists of a community, however, with different points of origins.

1.2. Beyond an Economic Rationale for Communities

In identifying interests, the NPD Report places predominant emphasis on 'unlocking economic benefit from non-personal data for India and its people', and the benefits that can accrue with the commodification of data.[17] This is a limited imagination of how communities may conceive their relationship with data, which in addition to being beneficial for market opportunities, also concerns people's cultural practices, privacy rights, as well community norms and interests. For example, predictive policing policies based on historical crime data may unfairly prejudice racial or religious minorities, or unfairly target certain residential localities. [18]Such communities therefore have a legitimate interest in preventing deployment of such technology. The NPD Report does not justify or explain why economic rights must be privileged over other legitimate interests. In fact, in a landmark decision affirming the fundamental right to privacy the Supreme Court has unequivocally rejected the idea that civil and political rights must yield to social-economic interests or

17 Committee of Experts (n 9) 6.

18 Vidushi Marda and Shivangi Narayan, 'Data in New Delhi's Predictive Policing System', *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (ACM 2020).

benefits.[19] The Court observed that strong civil and political rights or freedoms were essential to create conditions for achieving social progress, by creating a more just and empowered setting for interrogating the success of economic interventions.[20]

This limited economic standpoint towards safeguarding community rights suggests that the underlying cause of ensuring community agency is the extractive potential of data. This data capitalism under which interests of communities are also subsumed does not account for the inherent power disparities in our societies whether based on gender, caste, class, region etc.[21] These disparities will inevitably affect how communities are able to organise, speak, and have agency for their own welfare and development. In our view, by looking at identity which can be based on religion, region, caste, or class, we are looking at systemic structures that underpin how communities exist, and flourish, albeit not necessarily communities that people have consciously become a part of with a common purpose. This recognises that communities can be spaces that people inherit and feel familiar with but also spaces where society labels and conditions you to represent. Looking at regions within India for instance, we will also be able to distinguish how and why certain groups have less access and agency than others depending on the political, social, and economic conditions. For instance, internet shutdowns are commonplace in Jammu and Kashmir than in other parts of the country and as a result limit the agency of groups of students or businesses purely by virtue of their regional identity.

Further, in terms of interest we need to be able to expand what this might entail even beyond commercial interests as communities can be formed based on needs (for education, health), work (unions, informal workers), social causes (environmental justice, economic, political justice) and hence are both dynamic and reflexive of societal circumstance.

1.2.1. Internal dynamics of a community

Another area of concern in the recognition of Data Communities is that only those that are registered as companies under Section 8 of the Companies Act, 2013 (Companies Act), trusts or societies can raise complaints on behalf of the community.[22] This formalisation of how complaints are recognised for the purpose of grievance redressal is important, because it recognises only those

19 *Justice K.S. Puttaswamy (Retd). v. Union of India And Ors.* (2017) 10 SCC 1; Para 266-267.

20 Ibid.

21 Ameya Bokil and others, 'Settled Habits, New Tricks: Casteist Policing Meets Big Tech in India' (*Longreads*) <https://longreads.tni.org/stateofpower/settled-habits-new-tricks-casteist-policing-meets-big-tech-in-india> accessed 16 November 2022; Yeshimabeit Milner and Amy Traub, 'Data Capitalism and Algorithmic Racism' (Data for Black Lives and Demos 2021) <https://www.demos.org/research/data-capitalism-and-algorithmic-racism> accessed 16 November 2022.

22 Committee of Experts (n 9) 16.

data trustees that already have legal recognition, as mandated by the state. The existence of communities that might be algorithmically created by analysing group preferences (such as users of an online shopping platform) would not be able to respond until and unless they have trustees that have formal sanction. Further, many communities based on religion, tradition, custom who have existed for years, but without statutory recognition of their representatives, would be excluded by this categorisation. Instead, it might be more fruitful to recognise the collective nature of harms, and the collective rights that people have. This would address the real danger where communities are only those that receive official state sanction.

In thinking about redressal, we also need to acknowledge that grievances apply at multiple levels within a community and outside. At one level, it is important to distinguish individual interest, community interest, and public interest, and be careful not to conflate the interests of the community with individual interest, nor with public interest.[23] In certain circumstances, community interests may also appear to be parochial. For instance, residents of a locality where most homes have multiple personal vehicles might be opposed to development of public transport infrastructure near that locality. Therefore, while these interests may not be mutually exclusive, they are also nuanced, and there may be intersectional experiences when one accounts for distinctions based on gender, age, religion within different communities between individuals and communities. We need to account for different types of communities and examine how these intersect or conflict with the public interest and common good.

Therefore, while the broad definition proposed in the NPD Report is welcome, there is a need to clearly map out how group – or community – interests and identity are increasingly impacted by data processing and algorithmic decision-making. It is also important to formally recognise and include communities which may come to be formed solely because of common or shared interests. In addition to identifying such interests, we have also tried to argue that the primary basis for recognising a community should not be limited to its ability to extract commercial value out of its data, but must include other legitimate social, cultural, and political identity which a community may have an interest in protecting. In doing so, we have provided markers over how to think about the concept of a community, but do not offer a precise definition, as by design, we acknowledge the dynamic nature of communities.

---

23 Here is where Hess and Ostrom's distinction between different kinds of goods is valuable. For whereas community resources might be public resources where exclusions is difficult and subtractability is low (where one's use can reduce availability of others), they could also be common pool resources where exclusions remain difficult but subtractability is high, or clubs where both exclusions are easy to make, and subtractability is low or private goods where both exclusions are high and subtractability is high. Charlotte Hess and Elinor Ostrom, 'Ideas, Artifacts, and Facilities: Information as a Common-Pool Resource' (2003) 66 Law and Contemporary Problems 111 <https://scholarship.law.duke.edu/lcp/vol66/iss1/5> accessed 14 November 2022.

2. Community Data

2.1. The Emergence of Community Data

In our earlier discussion, we spoke of the fluid nature of communities, and the importance of thinking of them across a taxonomy of identity and interest. In this section, extending that argument, we explore how such fluidity intersects when thinking in terms of the ways in which communities organise and are shaped around data.

Community data as a concept emerges not just in terms of how communities may seek to govern the data that they generate when there is a shared and collective interest – for instance in terms of work or political activism. It also includes, as we have described earlier, communities that are formed based on algorithmic classifications when the data generated creates patterns and categories of groups of people.[24] In this second instance, the group is created on an ad hoc basis. For example, it could be in relation to shopping patterns, or social media interactions. Acknowledging the duality in the relation between community and data is important because it recognizes that often the formation of such groups is inherently by circumstance –they are not deliberate nor created with any agency from the individual. Therefore, by community data, we mean the nature or categories of non-personal data over which a Data Community may be able to exercise rights (Community Data).

2.1.1. Identifying Community Data

As with the definition of community, the policy documents referred to earlier also do not identify a clear basis for identifying or defining Community Data. The earlier version of the NPD Report attempted to provide a definition for Community Data by classifying non – personal data into three categories, 'Public Non-Personal Data', 'Community Non-Personal Data' and 'Private Non-Personal Data'. Public Non-Personal Data was defined as non – personal data collected by the state, or in the course of any publicly funded activity, such as census data.[25] Private Non-Personal Data included data collected by private entities, including inferred or derived data, such as customer surveys and other derived data/ insights inferred by applying algorithms.[26] Community Non-Personal Data on the other hand was classified as any kind of non-personal data whose 'source or subject pertains to a community of natural persons', such as

24 Brent Mittelstadt, 'From Individual to Group Privacy in Big Data Analytics' (2017) 30 Philosophy & Technology 475 <http://link.springer.com/10.1007/s13347-017-0253-7> accessed 16 November 2022.

25 Committee of Experts (n 8) 14.

26 ibid 15.

telecom or e-commerce data. However, Community Data specifically excluded all derived or processed data as well as all Private Non-Personal Data.[27]

In the subsequently released NPD Report, the Committee omitted this classification as well as any definition for Community Data. Instead, it limits the scope of a community's right to certain 'High Value Datasets' (HVD), which have been defined as 'a dataset that is beneficial to the community at large and shared as a public good, subject to certain guidelines...'.[28] The NPD Report clarifies that this could be a dataset deemed useful for policy making, improving public services, helping create new jobs or businesses, or for financial inclusion, poverty alleviation etc.[29]

The Committee's reconsideration of express exclusion of all privately collected non-personal data from the scope of Community Non-Personal Data (or Community Data for the purposes of this essay) is a welcome change. Given that the concept of community rights over data has evolved as an extension of the existing privacy and data governance framework,[30] exclusion of such data would have implied that Data Communities such as riders of food delivery platforms would not be able to exercise any rights over their Community Data, despite such data being used to determine their working hours, incentives etc. As set out above, while HVDs have been defined in very broad terms under the revised NPD Report, it remains unclear whether such data will be considered as a HVD since its purpose may be quite different from simply improving public services or job creation etc.

Similarly, while the Committee has dropped the reference to express exclusion of derived or aggregated data, it is not clear if the Committee now favours its inclusion within the scope of Community Data. Since aggregation of non-personal data allows entities (private and state) to gain newer insights about a group or community, limiting Community Data to only 'raw' or 'factual data' would not address the vacuum created by existing data protection or governance frameworks. On the contrary, it would only exacerbate the existing information asymmetry between Data Communities (such as drivers associated with ride-hailing platforms) and the holders of Community Data (such as ride-hailing platforms).

---

27 While the definition excludes Private Non - Personal Data, the illustrations set out in the Report indicate that even data held by private telecom companies and ride hailing companies be included, thereby suggesting that the Committee did not intend to perhaps exclude all Private Non Personal Data.

28 Committee of Experts (n 9).

29 ibid 18.

30 Committee of Experts under the Chairmanship of Justice B. N. Srikrishna (n 5) 46–46.

Therefore, it is imperative that the guidelines proposed to be framed for identifying HVDs expressly include datasets containing privately held non-personal data, as well as derived or aggregated data, subject to claims of copyright or trade secrets that may legitimately be invoked. However, as argued above, commercial exploitation is only one among several interests that a framework for Community Data must aim to secure. If Community Data is being sought to address issues of information asymmetry and preventing discriminatory outcomes, it may be worthwhile to explore furthering fair use exceptions to intellectual property claims. This is particularly relevant when we are thinking of questions of data justice from the standpoint of communities. This would involve investigating the ways in which people are made visible, represented, and treated as a result of the production of their data.[31] In thinking about questions of justice within community data, we want to recognise the political economy that underpins a digital economy, and the hierarchies of relations, institutions, that might create marginalisation within members of a community.[32] Therefore in thinking about community benefit and public good, as in the HVD, we also need to account for the internal dynamics that exist within communities, and the need to ensure that these are accounted for.

2.1.2. Ensuring access to data for communities

We, therefore, believe that non-personal data should not be classified based on who has collected it, as such classification is immaterial to whether the data itself is valuable to the community or not. Instead, regulators may consider a classification based on the nature of data collected i.e., whether it is anonymised personal data, or data which was originally not personal data, (such as data related to wind, climate, agricultural produce etc.). Communities may access either kind of non-personal data, but the degree of access to, and the rights that may be exercised over, the two kinds of data may need separate approaches based on the risks involved.

For example, with respect to anonymised personal data, given the widely acknowledged risks associated with its de-anonymization,[33] a graded approach may be warranted when facilitating access to it to communities. As suggested in the NPD Report itself, such a graded approach may be based on the underlying 'sensitivity' of the personal data which is anonymised, such as when the

---

31 Linnet Taylor, 'What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally' (2017) 4 Big Data & Society 2053951717736335 <http://journals.sagepub.com/doi/10.1177/2053951717736335> accessed 16 November 2022.

32 Barbara Prainsack, 'The Political Economy of Digital Data: Introduction to the Special Issue' (2020) 41 Policy Studies 439 <https://www.tandfonline.com/doi/full/10.1080/01442872.2020.1723519> accessed 16 November 2022.

33 Boris Lubarsky, 'Re-Identification of "Anonymised Data"' (2017) 1 Georgetown Law Technology Review 202 <https://georgetownlawtechreview.org/re-identification-of-anonymized-data/GLTR-04-2017/>; C. Christine Porter, 'De-Identified Data and Third Party Data Mining: The Risk of Re-Identification of Personal Information' (2008) 5 Shidler Journal of Law, Commerce +Technology 3 <https://digitalcommons.law.uw.edu/wjlta/vol5/iss1/3> accessed 16 November 2022.

underlying data is health or financial data.[34] This distinction is also legitimate in view of the stricter controls over sensitive personal data envisaged under most data protection legislations[35] and in order to reduce the risk of individual privacy rights being compromised while facilitating access to Community Data. A graded approach can similarly also be employed in the context of derived or aggregated data, with a community having limited rights over such data, such as the rights to access and to object to processing of data in a particular manner subject to establishing injury or harm to the community. This is akin to an individual's right to access and object to the processing of personal data, especially automated data, as guaranteed under some data protection legislations.[36]

Further, as already discussed in the previous section, while the NPD Report acknowledges collective privacy harms arising out of sophisticated data analytic techniques, and recognises that such data may have a bearing on the well-being, rights and dignity of the community, the framing appears to place less importance on these concerns when compared to the emphasis of extracting commercial value from data.[37] To this end, we propose that in addition to the criteria already identified by the Committee for determining a HDV, objectives such as reducing information asymmetry, challenging discriminatory outcomes, and seeking better working conditions/ fairer workplace policies may also be added. In this regard, we recommend the inclusion of private and aggregated data, as well as the classification of data based on the source, thereby broadening the criteria of HVDs.

3. Community Data Governance

3.1. The Role of Governance

How do these arguments on community and community data connect to governance? Governance as a term takes on several meanings. It can refer to the management through which a system is controlled, a set of rules and procedures that mandate behaviours. It also has a normative scope which consists of principles such as accountability, transparency, increased public participation that can create value and ground decision making processes.[38] In our view, taking a more normative look at governance is important because

34 Committee of Experts (n 9) 16.

35 Regulation (Eu) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679> accessed 17 November 2022; The Personal Data Protection Bill, 2019, Lok Sabha (Bill No. 373 of 2019)' <http://164.100.47.4/BillsTexts/LSBillTexts/Asintroduced/373_2019_LS_Eng.pdf> accessed 17 November 2022.

36 Ibid, Articles 15, 21.

37 Committee of Experts (n 9) 25.

38 Marina Micheli and others, 'Emerging Models of Data Governance in the Age of Datafication' (2020) 7 Big Data & Society 2053951720948087 <http://journals.sagepub.com/doi/10.1177/2053951720948087> accessed 17 November 2022.

doing so will allow for a more comprehensive examination of how governance frameworks are designed.[39] This would include examining not just principles for data sharing; the ways in which data is used but also the opportunity that individuals and groups have to access and participate in data sharing. It would also include thinking through the intentions of a governance framework from enhancing economic growth to ensuring public interest, to encouraging private enterprise while setting out the terms for how this can be achieved.[40]

Underlying the different forms of data governance are two large trends. The first is an emphasis on thinking about ways to maximise the value of data; and the second is the ways in which data can be used to solve problems in contextually determined ways.[41] In each of these considerations, however, it is important to think about the politics that underlie data governance frameworks. For instance, as indigenous data governance activists have advanced, open data movements of FAIR principles for findable, accessible, interoperable, and reusable data place a pre-eminence on data sharing. However, these approaches do not account for power differentials that exist between individuals and groups. Hence, they argue that it is important to also have complementary principles which advocate for principles around Collective Benefit, Authority to Control, Responsibility and Ethics.[42] This shift is grounded in a need to be able to center people, and their contexts in terms of how data is governed. It is to ensure that there is emphasis on not just greater participation but also more equitable outcomes and underpins our understanding of how data should be governed.[43]

Our understanding of community data governance flows from the unpacking of community, and Community Data. We believe that governance frameworks should reflect the heterogeneity in terms of how communities are formed, how they operate and then disappear, as well as the ways that members relate to the data about them. This means that members should have agency to understand what and where the data is used, how it used, as well as how it impacts them.[44]

---

39 European Parliament. Directorate General for Parliamentary Research Services., *Governing Data and Artificial Intelligence for All: Models for Sustainable and Just Data Governance*. (Publications Office 2022) <https://data.europa.eu/doi/10.2861/915401> accessed 17 November 2022.

40 Aditi Ramesh, 'Community Data Governance and Its Application for Migrant Communities in Urban India' (The Data Economy Lab 2020) <https://thedataeconomylab.com/2020/10/29/community-data-governance-and-its-application-for-migrant-communities-in-urban-india/> accessed 17 November 2022.

41 Sean Martin McDonald, 'Data Governance's New Clothes' (*Centre for International Governance Innovation*) <https://www.cigionline.org/articles/data-governances-new-clothes/> accessed 17 November 2022.

42 'CARE Principles of Indigenous Data Governance' (*Global Indigenous Data Alliance*) <https://www.gida-global.org/care> accessed 17 November 2022. Collective benefit- "Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data"; Authority to control- "Indigenous Peoples' rights and interests in Indigenous data must be recognised and their authority to control such data be empowered"; Responsibility- "Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples' self-determination and collective benefits" and Ethics - "Indigenous Peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem".

43 Stephanie Russo Carroll and others, 'The CARE Principles for Indigenous Data Governance' (2020) 19 Data Science Journal 43 <http://datascience.codata.org/articles/10.5334/dsj-2020-043/> accessed 4 November 2022.

44 Taylor (n 30).

Doing so acknowledges that governance frameworks consider communities that are fluid by design and that the frameworks do not constrain their capacity to organize and thrive through needless formalization.

3.1.1. Role of representatives

The NPD Report proposes that a community exercise its rights through a data trustee.[45] While the original NPD Report did not define a data trustee, the Revised NPD Report defines a data trustee as an organisation (either government or non-profit private entity such as a company incorporated under Section 8 of the Companies Act, a society or a trust) that is responsible for the creation, maintenance, sharing of certain HVDs.[46] The Revised NPD Report further clarifies that a specialised regulator for non-personal data will have to issue guidelines to determine the appropriateness of a given data trustee, who in-turn has the power to request the regulator to create a HVD for and on behalf of the community. The community itself may access the HVD by approaching the data trustee, who owes a 'duty of care' to the community, an obligation to ensure that no member(s) of the community are harmed by re-identification of the data. The trustee is also required to establish a grievance redressal mechanism.[47]

We welcome the Committee's recommendation that in principle, a representative of the community itself must identify and suggest which datasets may be valuable to the community it represents. As mentioned above, the Revised NPD Report does not provide any justification for envisaging a data trustee to be only a government agency or a non-profit private organization. It does not require that at least, to the extent possible, the data trustee be an individual belonging to the community in question, or an entity composed of members from that community. Under other legal frameworks governing community rights, 'trustees' or agents of the community are often elected representatives of the community,[48] thereby ensuring some degree of representation. Similarly, appointing government entities as data trustees raises important issues of conflict of interest if the government itself is the regulator, and the custodian of data in each case.[49] In cases where appointing a government entity is inevitable, additional safeguards may be required to avoid conflicts of interest

45 Committee of Experts (n 9).

46 ibid 18.

47 ibid.

48 Puneeth Nagaraj, Varsha Rao, and Dedipyaman Shukla, 'Community Rights Over Non-Personal Data: Perspectives from Jurisprudence on Natural Resources' (Data Governance Network) <https://datagovernance.org/files/research/1611826214.pdf> accessed 17 November 2022.

49 Centre for Communication Governance, NLU Delhi, 'Comments to MEITY on the Report by the Committee of Experts on Non-Personal Data Governance Framework.' (2020) <https://ccgdelhi.org/wp-content/uploads/2020/09/CCG-NLU-Comments-to-MeitY-on-the-Report-by-the-Committee-of-Experts-on-Non-Personal-Data-Governance-Framework.pdf> accessed 19 October 2021.

such as ensuring that the interests of the appointing authority and the trustee are clearly articulated, ensuring that officials responsible document and are able to justify their decisions as appropriate to public scrutiny, and ensuring that the trustees are responsive to the needs of the community.

The recent criticism of public interest litigations has important lessons to offer in this context. In the 1980s, in an attempt to remove barriers to access to justice, and to democratise judicial remedies,[50] the Supreme Court diluted the tests for *locus standi* (the petitioner's standing) for public interest litigations and held that it would, in a fit case, take cognizance of petitions preferred by 'public spirited individuals', who were not directly affected by the action complained of.[51] Anuj Bhuwania has pointed out how this supposed procedural informality has often resulted in the Court passing directions without actually hearing any of the affected parties.[52] Consequently, several scholars have suggested that the Supreme Court revisit this judicial 'innovation' and adopt a more principled approach to public interest litigations.[53] A better way forward for operationalising data trusts, and facilitating sharing of Community Data would perhaps be to adopt the test contemplated under the Civil Procedure Code, 1908 (CPC). The CPC contemplates an individual bringing a claim in representative capacity.[54] Interpreting this provision, the Supreme Court has held that while granting leave to permit institution of such suits, Courts only need to look at whether the persons on whose behalf the action is being brought have the same interest.[55] Clarifying this further, the Court held that either the interest must be common, or all persons must have a common grievance which they seek to get redressed. While such a requirement may make it harder to appoint a data trust, it may ultimately lead to the appointment of a more representative agent on behalf of a given community.

3.1.2. Governance and harms

Data governance models needs to be able to account for the ways in which the production of data causes 'harms' arising out of algorithmic decision-making or profiling which sometimes may be remote and not easily identifiable.[56] For community governance to be able to speak to the collective benefit of the

---

50 *People's Union for Democratic Rights v. Union of India*, (1982) 3 SCC 235.

51 Anuj Bhuwania, *Courting the People: Public Interest Litigation in Post-Emergency India* (1st edn, Cambridge University Press 2016) 30 <https://www.cambridge.org/core/product/identifier/9781316551745/type/book> accessed 17 November 2022.

52 ibid 82–120.

53 ibid 120.

54 Order 1 Rule 8, Code of Civil Procedure, 1908.

55 *Chairman Tamil Nadu Housing Board v. T.N Ganapathy* (1990) 1 SCC 608.

56 Danielle Keats Citron and Daniel J Solove, 'Privacy Harms' [2021] SSRN Electronic Journal <https://www.ssrn.com/abstract=3782222> accessed 17 November 2022.

represented community, where interests are not only represented but also responded to, a wider definition of possible harms may be required to facilitate meeting this threshold of 'common interest' or a 'common grievance'. In a recent paper, Danielle Citron and Daniel Solove observe:[57]

"Courts struggle with privacy harms because they often involve future uses of personal data that vary widely. When privacy violations do result in negative consequences, the effects are often small – frustration, aggravation, and inconvenience – and dispersed among a large number of people. When these minor harms are done at a vast scale by a large number of actors, they aggregate into more significant harms to people and society. But these harms do not fit well with existing judicial understandings of harm."

Based on this, Citron and Solove set out a typology of harms which ought to be recognised by courts. These are: physical harms, economic harms (financial loss as well as the loss of economic opportunities), reputational harms, emotional harms, relationship harms (arising out of loss of confidentiality and damage to trust), chilling effect harms, discrimination harms, thwarted expectation harms (improper use of data), control harms, data quality harms (harms arising out of inaccurate or incorrect data), informed choice harms, vulnerability harms (harms arising out of failing to secure data properly), disturbance harms and autonomy harms (when data is used to manipulate individuals and coerce / suggest outcomes).[58] While argued in the context of personal data and individual rights, we believe that the typology of harms identified offers a useful framing for understanding and articulating collective harms as well, which may be experienced by communities owing to the extensive aggregation of non-personal data. This typology of harms can be appropriately modified from the point of view of a community to include and address collective harms, making it easier for communities to identify which data may be valuable, and exercising rights over it.

Conclusion

In this essay, we discussed the idea of Community Data that was introduced in the Non-Personal Data framework. We aimed to engage with existing commentaries on the definitional challenges around who is a community, how it is constituted, who it represents, as well as propose a framework to be able to explore how to operationalize the concept.

57 ibid.

58 ibid.

To do this, we developed three concepts in this essay: community, community data and community data governance. In our understanding, while these concepts are necessarily interwoven and connected, they still require to be distinguished to be able to explore how to think of community data from a regulatory perspective.

Our approach to thinking about community has involved examining how to incorporate a fluid idea of community which may emerge based on identity and interest; as well as community data, which may restrict the agency of the community that may also be generated through algorithmic classifications. In doing so, we are interested in demonstrating how such a concept can bring representation for new groups, coalitions, and alliances as a by-product of participating in a digital economy, for instance worker unions of platform workers – in addition to acknowledge existing coalitions. We have argued for thinking of group rights, while remaining mindful of hierarchies within groups and the rights of individuals within a group. Finally, we have identified the interests or outcomes that any framework for community data must - at the very least - aim to secure and the possible regulatory frameworks for articulating or securing such outcomes.

For What it's Worth: Realising the Value of Data

Mansi Kedia and Gangesh Varma[1]

Introduction

Economic resources, human or other, utilised to produce output were referred to by early economists as factors of production. From the classic set of four - land, labour, capital, and entrepreneurship, the factors were expanded to include other natural resources, raw materials and also information.[2] Information has been defined by some as data affecting behaviour.[3] The role of information (*processed data*)[4] as a distinct factor of production was recognised several decades ago and was tested using guiding principles of factor markets - factor prices, sources of supply, sources of demand and ability to produce additional products.[5] The economic value of data is thus apparent. However, the imagination and measurement of its scale and scope have been completely transformed by the digital economy. While we may have witnessed several socio-economic transitions in the past, the current digitalisation is an unimaginable scale of civilisational transformation, with data at its centre.

Data is the "raw material produced by abstracting the world into categories, measures and other representational forms, such as – numbers, characters, symbols, images, sounds, electromagnetic waves, bits, etc.".[6] The exponential growth of data and its processing has led to ubiquitous use cases. Companies collect and use data for innovation, process efficiency, marketing, and customization of services. Governments rely on data to improve governance, quality and reach of public services that it renders. These applications cut across a range of data types - demographic data, personal data, spatial data, machine data, etc. However, the production and utilisation of data is not new. The history of such utility ranges from scientific innovations to state policy, notwithstanding the time and cost invested in generating, analysing, and interpreting it. It wouldn't be an exaggeration to say

---

1 Mansi is Senior Fellow at the Indian Council for Research on International Economic Relations. She can be reached at mkedia@icrier.res.in. Gangesh is a Senior Associate at Saraf and Partners focusing on technology and policy. The author can be reached at gangeshvarma@gmail.com.

2 Sunday Okpighe, 'The Seven Factors of Production' (2015) 5 British Journal of Applied Science & Technology 217 <https://journalcjast.com/index.php/CJAST/article/view/510> accessed 12 November 2022.

3 A. M. McDonough, *Information Economics and Management Systems* (McGraw Hill 1963).

4 When data is processed, organized, structured, or presented in a given context so as to make it useful, it is called information.

5 Andrew Berczi, 'Information As A Factor Of Production' (1981) 16 Business Economics 14 <http://www.jstor.org/stable/23482505> accessed 12 November 2022.

6 Rob Kitchin, *The Data Revolution* (SAGE Publishing 2014).

that good-quality data was a scarce resource, and therefore treated as a valuable commodity, which was either carefully shielded or traded.[7]

With the advent of the internet of things (IoT), the sources and types of data have grown exponentially. In a world of hyper digital connectivity, numerous tools, and devices, ranging from watches to washing machines are equipped to become a source of data. This paradigmatic shift is reflected in how data is produced, collected, stored, and utilised. From being scarce and limited in access, today we have multiples of quintillion (*1 followed by 18 zeros*) bytes of data being generated every day. It is less costly, high frequency, targeted and much more accurate. It is also relatively open and accessible. As Rob Kitchn stated in *The Data Revolution*, "A data revolution is underway, one that is already reshaping how knowledge is produced, business is conducted, and governance is enacted."[8]

The enormous volume of data is aptly captured by the nomenclature 'Big Data' which requires greater analytical and processing capabilities in contrast to the earlier treatment of data. Consequently, its applications are much more complex and demanding. However, big data is not the only component of the data revolution. Related initiatives include digitisation, linking together, and scaling-up of traditionally produced datasets (small data) into networked data infrastructures, and developing new indicators, targets or open datasets that has directly fed into the development discourse around the world.[9]

The nature and applications of data have led to the creation of institutions that attempt to standardize guidelines and policies with respect to data formats, sharing protocols and intellectual property rights regimes. The nature and function of these institutions are very diverse. For example, institutions like the Internet Engineering Task Force, founded in 1986, develop protocols and standards for the internet and that, to an extent, determines the treatment of data on the internet as a medium. The World Intellectual Property Organisation in 1997, adopted guidelines for the protection of databases. More recently, countries and regions have created their own regulatory frameworks such as the APEC Privacy Framework, the European Union's General Data Protection Regulation, and the ASEAN Framework on Personal Data Protection. Prima facie, most current institutions and regulations consider trade-offs and prioritize one of the many aspects of data governance. The introduction to this essay might suggest that the value of data is only manifested in the economic outcomes it catalyses, a representation of monetary benefit or the importance of data in reducing costs, improving efficiency, scaling up and other factors relating to

---

7 ibid.

8 ibid.

9 Independent Expert Advisory Group on Data Revolution for Sustainable Development, 'A World That Counts: Mobilising the Data Revolution for Sustainable Development' (2014) <https://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf> accessed 12 November 2022; World Bank, 'The World Bank Group Supports the Data Revolution for Sustainable Development' <https://www.worldbank.org/content/dam/Worldbank/Statcap/HDRSD/WBG-support_data_revolution.pdf> accessed 14 November 2022.

economic growth.[10] In our assessment, as reflected by others[11] this is a narrow view of the value of data and we propose that a comprehensive framework that looks at its social and technical aspects be included in the assessment of the value of data. A paper by Gunther et al (2017) recommends an integrated model for realising the value of data for individual organizations, though social well-being is measured only in terms of education, health, public safety, and security.[12] In this essay, we propose to strengthen this idea and offer a framework for a comprehensive measurement of the value of data that includes its economic, social and technical aspects. The future of the digital economy will be better understood if we are able to measure it.

The following sections of the essay will outline the proposed framework, explaining different kinds of value drivers and their implications on data governance. It also provides an example to illustrate the integration of value drivers and an implementation model for data governance.

1. A Framework for the Value of Data and Implications on Data Governance

The understanding of the value of data has evolved over time, adapting to its ever-increasing applications. The current discourse on data governance mainly adopts two types of value frameworks, one that focuses on pure profit (economic value generation) and the other on human rights.[13] Those highlighting economic profit focus on the unfettered use of data for improving the quality and reliability of business processes to maximise monetary gains, while champions of human rights emphasize on the harms of data abuse or misuse, prioritising user rights and opposing its ungoverned utilisation. However, in our opinion, this binary approach to assessing the value of data can result in a rather polarizing set of perspectives on data governance and policy development.[14]

---

10 Andrew McAfee and Erik Brynjolfsson, 'Big Data: The Management Revolution' [2012] *Harvard Business Review* <https://hbr.org/2012/10/big-data-the-management-revolution> accessed 12 November 2022.

11 James Wilson and others, 'The Value of Data: Applying a Public Value Model to the English National Health Service' (2020) 22 Journal of Medical Internet Research e15816 <http://www.jmir.org/2020/3/e15816/> accessed 14 November 2022; Jaap Wieringa and others, 'Data Analytics in a Privacy-Concerned World' (2021) 122 Journal of Business Research 915 <https://linkinghub.elsevier.com/retrieve/pii/S0148296319303078> accessed 14 November 2022.

12 Wendy Arianne Günther and others, 'Debating Big Data: A Literature Review on Realizing Value from Big Data' (2017) 26 The Journal of Strategic Information Systems 191 <https://linkinghub.elsevier.com/retrieve/pii/S0963868717302615> accessed 14 November 2022.

13 Bill Schmarzo and Dr. Mouwafac Sidaoui, 'Applying Economic Concepts To Big Data To Determine The Financial Value Of The Organization's Data And Analytics, And Understanding The Ramifications On The Organizations' Financial Statements And IT Operations And Business Strategies' (2017) <https://www.dell.com/wp-uploads/2017/03/USF-The-Economics-of-Data-and-Analytics-Final2.pdf>; HM Treasury, 'The Economic Value of Data: Discussion Paper' (2018) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/731349/20180730_HMT_Discussion_Paper_-_The_Economic_Value_of_Data.pdf>.

14 Ivana Kottasová, 'These Companies Are Getting Killed by GDPR' (*CNNMoney*, 11 May 2018) <https://money.cnn.com/2018/05/11/technology/gdpr-tech-companies-losers/index.html> accessed 14 November 2022 (several companies have had to shut down operations due to GDPR); Rebecca Janßen and others, 'GDPR and the Lost Generation of Innovative Apps' (National Bureau of Economic Research, May 2022) <https://www.nber.org/papers/w30028> accessed 14 November 2022 (These findings which sate that the GDPR has has an adverse impact on innovation should not be overstated); Joseph Jerome, 'The GDPR's Impact on Innovation Should Not Be Overstated' (*Center for Democracy and Technology*) <https://cdt.org/insights/the-gdprs-impact-on-innovation-should-not-be-overstated/> accessed 12 November 2022; Nicholas Martin and others, 'How Data Protection Regulation Affects Startup Innovation' (2019) 21 Information Systems Frontiers 1307 <http://link.springer.com/10.1007/s10796-019-09974-2> accessed 14 November 2022. (For a mixed perspective where research results show both the stimulation and stifling of innovation).

Our purpose here is to illustrate the need for a comprehensive value framework that helps harness the potential of data without causing harm and minimising abuse. This idea is not new but needs reiteration and mainstreaming for policymakers to recognise its importance in designing data governance policies. We examine conceptions of both data and value.

While the definition of data was introduced earlier in this essay, its varied conceptions have not been explored. There are several metaphors and analogies used to describe data and make sense of it. Comparisons range from oil[15] to sunshine,[16] avocado[17] to nuclear power,[18] but these rarely offer consolation to those trying to make sense of the role and value of data. While there may be several more analogies that may be developed to explain the exploitation of data, each one becomes inadequate in a fresh context. However, these diverse conceptions reflect on its extensive utility and value drivers. From the economics point of view, the value of data can be paralleled to its conceptualisation as capital,[19] infrastructure,[20] currency,[21] asset[22] or generally as an economic good. Arguing that it is non-rivalrous and non-excludable, data is predominantly considered as a public good that should be made accessible and used to deliver real value to society.[23]

Greater utilisation of data and demands for regulating its use led to the emergence of privacy guidelines that offered multiple taxonomies for data. With digitalisation there is a fundamental change in the data itself. Data that originates from observations today are less obvious to the individual and are a product of processing itself.[24] More recently, regulators have delineated data into personal data and non-

15 Charles Arthur, 'Tech Giants May Be Huge, but Nothing Matches Big Data' *The Guardian* (23 August 2013) <https://www.theguardian.com/technology/2013/aug/23/tech-giants-data> accessed 14 November 2022.

16 Shona Ghosh and Jake Kanter, 'Google Says Data Is More like Sunlight than Oil, Just 1 Day after Being Fined $57 Million over Its Privacy and Consent Practices' (*Business Insider*, 22 January 2019) <https://www.businessinsider.in/google-says-data-is-more-like-sunlight-than-oil-just-1-day-after-being-fined-57-million-over-its-privacy-and-consent-practices/articleshow/67640224.cms> accessed 12 November 2022.

17 Dr. Deborah Elms, 'Data Is the New Avocado?' (*Asian Trade Centre*, 9 April 2019) <https://asiantradecentre.org/talkingtrade//data-is-the-new-avocado> accessed 12 November 2022.

18 James Bridle, 'Opinion: Data Isn't the New Oil — It's the New Nuclear Power' (*ideas.ted.com*, 17 July 2018) <https://ideas.ted.com/opinion-data-isnt-the-new-oil-its-the-new-nuclear-power/> accessed 12 November 2022.

19 Jathan Sadowski, 'When Data Is Capital: Datafication, Accumulation, and Extraction' (2019) 6 Big Data & Society 2053951718820549 <http://journals.sagepub.com/doi/10.1177/2053951718820549> accessed 14 November 2022.

20 Peter Kawalek and Ali Bayat, 'Data as Infrastructure' <https://nic.org.uk/app/uploads//Data-As-Infrastructure.pdf> accessed 12 November 2022.

21 Knowledge at Wharton Staff, 'Data as Currency: What Value Are You Getting?' (*Knowledge at Wharton*) <https://knowledge.wharton.upenn.edu/article/barrett-data-as-currency/> accessed 14 November 2022; William D Eggers, Rob Hamill and Abed Ali, 'Data as the New Currency: Government's Role in Facilitating the Exchange' [2013] Delloit Review <https://www2.deloitte.com/content/dam/insights/us/articles/data-as-the-new-currency/DR13_data_as_the_new_currency2.pdf> accessed 12 November 2022; Guillaume Desjardins, 'Your Personal Data Is the Currency of the Digital Age' (*The Conversation*) <http://theconversation.com/your-personal-data-is-the-currency-of-the-digital-age-146386> accessed 14 November 2022.

22 Peter Lake and Paul Crowther, 'Data, an Organisational Asset' in Peter Lake and Paul Crowther, *Concise Guide to Databases* (Springer London 2013) <http://link.springer.com/10.1007/978-1-4471-5601-7_1> accessed 12 November 2022.

23 'The Data Centered Economy: A New Temple for a New India' (Indian Council for Research on International Economic Relations 2019) <http://icrier.org/pdf/The-Data-Centred-Economy-A-New-Temple-for-a-New-India.pdf> accessed 12 November 2022.

24 Martin Abrams, 'The Origins of Personal Data and Its Implications for Governance' [2014] SSRN Electronic Journal <http://www.ssrn.com/abstract=2510927> accessed 14 November 2022.

personal data. This was arguably a middle-path that protected individuals as well as facilitated innovation especially since it demarcated specific types of data and did not create a blanket restriction on the use of all types of data. Probably, the potential utility of non-personal data for developing both better products and better policies increased in its perceived value, and thereby brought more regulatory attention to the category of non-personal data. This scrutiny also highlights that there exists a certain degree of fluidity between these types of data where the anonmyisation of personal data could convert it to non-personal data whereby the restrictions could be diluted without threatening an individual's privacy. However, with the evolution of technology and data use, we are forced to question its efficacy.[25] Researchers have shown how they were able to de-anonymise data and render the protection provided by anonymisation pointless. Even with anonymisation the risk to privacy persists.[26] Regulations on data sharing infrastructure and data exchanges respond to the various types of data such as geo-spatial, financial, health, industrial, and demographic that are broadly categorised into personal and non-personal data. Examining different value dimensions of data is essential to illustrate the limitations of using an either-or approach.[27]

We believe that value from data can be classified into three main categories - social, economic, and technical. These can manifest in various permutations and combinations and could also result in overlaps. While notoriously difficult to measure, the concepts of economic and social value have often been discussed,[28] whereas technical value requires familiarization.

1.1. Economic Value of Data

As discussed above, data metaphors are often used to capture the economic value of data speaking of some part of its character, but rarely do they ever provide a complete picture. Data has distinctive traits such as its reusability. This also means that the often used "data is the new oil" is a poor metaphor. Its reusability and inexhaustible character renders it very different as a resource from that of a fossil fuel. Data can be reused endlessly and in domains like

25 Luc Rocher, Julien M Hendrickx and Yves-Alexandre de Montjoye, 'Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models' (2019) 10 Nature Communications 3069 <http://www.nature.com/articles/s41467-019-10933-3> accessed 14 November 2022.

26 Alex Hern, '"Anonymised" Data Can Never Be Totally Anonymous, Says Study' *The Guardian* (23 July 2019) <https://www.theguardian.com/technology/2019/jul/23/anonymised-data-never-be-anonymous-enough-study-finds> accessed 14 November 2022; Michele Finck and Frank Pallas, 'They Who Must Not Be Identified—Distinguishing Personal from Non-Personal Data under the GDPR' 10 International Data Privacy Law 11.

27 In some ways, the convergence of the regulatory scope on personal and non-personal data by the Joint Parliamentary Committee in its review of the Indian Data Protection Bill 2019 and then later the withdrawal of the bill by the government to reconsider the approach reflects the importance of carrying out a multi-dimensional value assessment at the pre-liminary stages of policy development.

28 Federico Cabitza, Angela Locoro and Carlo Batini, 'Making Open Data More Personal Through a Social Value Perspective: A Methodological Approach' (2020) 22 Information Systems Frontiers 131 <http://link.springer.com/10.1007/s10796-018-9854-7> accessed 14 November 2022; Günther and others (n 11); 'World Development Report 2021: Data for Better Lives' (The World Bank Group 2021) <https://www.worldbank.org/en/publication/wdr2021>; HM Treasury (n 12).

artificial intelligence and machine learning models, its value grows greater with scale. Data creates positive and negative externalities. These often cross-over to socio economic consequences. An individual's car journey when used with digital maps, saves them from traffic congestion, is an example of a positive externality. On the other hand, the use of an individual's personal data to exclude and discriminate against communities, especially those who did not share their data, is an example of a negative externality.

Mariana Mazzucato, professor in the economics of innovation and public value at University College London, thinks that profits in the digital technology era have become confused with value. She draws parallels to critiques of GDP as a misleading indicator when looked at through different lenses and asks important questions - What are the market participants doing? This is similar to the broader macroeconomic literature that challenges the use of GDP as a measure of welfare. Several empirical studies found that mean welfare stagnated and even deteriorated in many developed countries despite steady rise in GDP.[29] The parallel in the digital ecosystem could be an exponential increase in scale of digital adoption, accompanied by exploitative business models that profit a tiny group of organisations, while flawed algorithms and poor governance raise serious ethical concerns of consumer harm.[30] This highlights the limitation of using a single lens to measure the value of data.

The debate on the economic identity of data is also one that remains unsettled. As mentioned earlier, one of the various conceptions of data is as capital. Many businesses, especially in the digital ecosystem consider data as their single biggest asset. Proponents of this view regard data as a form of capital just like financial capital, given its ability to generate new products and services.[31] Data also creates new data. For those treating data as labour, the argument is centered around data as a measure to reduce inequality. Data intensive companies do not compensate consumers for their data, a resource that is monetised and exploited by businesses for financial gain.[32] The economic treatment of data varies with differing types of data and the purposes they serve. A cookie cutter approach in determining the economic value of data, without considering the context in which data is being collected and used, would be unfair and inefficient.

1.2. Social Value of Data

29 Jeroen Van den Bergh and Antal Miklós, *Evaluating Alternatives to GDP as Measures of Social Welfare/Progress* (Vienna : WWWforEurope 2014).

30 MIT Technology Review Insights, 'Fair Value? Fixing the Data Economy' (*MIT Technology Review*) <https://www.technologyreview.com/2020/12/03/1012797/fair-value-fixing-the-data-economy/> accessed 13 November 2022.

31 MIT Technology Review Custom in partnership with Oracle, 'The Rise of Data Capital' (2016) <http://files.technologyreview.com/whitepapers/MIT_Oracle+Report-The_Rise_of_Data_Capital.pdf> accessed 13 November 2022.

32 Wendy C.Y. Li, Makoto Nirei, and Kazufumi Yamana, 'Value of Data: There's No Such Thing As A Free Lunch in the Digital Economy' <https://www.oecd.org/site/stipatents/programme/ipsdm-2018-5-2-li-nirei-yamana.pdf> accessed 13 November 2022.

Social value of data arises from its ability to be utilised by and for communities or society at-large. The key distinction being that its use is not restricted to the economic considerations and benefits to an individual or singular entities in society. Open data sets are a great example of the potential social value of data beyond the economic gains it offers. Social value has been interpreted as the ability of data to be used for the benefit of and by the sources of data (citizens themselves).[33] This understanding is built around the public good argument and the non-rivalrous and non-excludable characteristics of data use. For instance, data on the health status of individuals in the pandemic or crowd sourced traffic updates create positive spill overs benefitting societies as a whole. This has featured in India's policy discourse at multiple stages, but most prominently as the discussion around 'community data' in the Report of the Expert Committee on Non-Personal Data which may be perceived as a call for 'distributive justice in a digital economy'.[34] The existence of informational externalities and the non-rival character of data immediately imply that private markets uses and market prices (if they exist), will not deliver social value. What's more, the value of any given data set is also fundamentally determined by the value of the uses to which it can be put, which are likely unknown until after the fact.[35] Yet for public controllers of data concerned to maximise social welfare, methods based on realised financial values in market transactions are insufficient. Social value could potentially be gained from more data collection, wider access, or the scope to join information from different data sets with varying types of data records (noting also the need to manage the negative externalities of potential privacy loss and security breaches).[36] In some sense social value of data is an aggregation of the positive and negative externalities, or the net social welfare.[37] A non-integrated approach can lead to sub-optimal outcomes as illustrated by Coyle & Diepeveen (2021) using the application of geospatial data and the transport sector in the UK.[38]

1.3. Technical Value of Data

According to MIT's recent Tech Review Insights, valuing data means understanding who participated in its creation. Data's value is also a product of the input and participation of digital users with complex consent protocols,

33 Cabitza, Locoro and Batini (n 27).

34 Aniruddh Nigam, '[Vidhispeaks] Exploring Community Data Rights over Non-Personal Data' (30 October 2021) <https://www.barandbench.com/columns/policy-columns/vidhispeaks-exploring-community-data-rights-over-non-personal-data> accessed 13 November 2022.

35 Diane Coyle and Annabel Manley, 'Working Paper - Potential Social Value from Data: An Application of Discrete Choice Analysis' (Bennett Institute for Public Policy, University of Cambridge, UK 2021) <https://www.bennettinstitute.cam.ac.uk/publications/social-value-data/> accessed 12 November 2022.

36 Diane Coyle and Stephanie Diepeveen, 'Creating and Governing Social Value from Data' [2021] SSRN Electronic Journal <https://www.ssrn.com/abstract=3973034> accessed 13 November 2022.

37 ibid.

38 ibid.

from granting permissions to platforms to access their data, to labelling and digitization of work conducted during processes like reCAPTCHA, etc.[39] We conceptualise technical value as the efficiency gains from the architecture of technology using which data is collected, processed, and stored. With dramatic increases in the volume of data structuring, data centre networks, and interconnecting architectures become critical to allow for the efficient use of data. Data management and data centre networks must choose from competing priorities routing efficiency, high capacity, low power consumption, flexibility, etc. Technical value is also grounded in the context that technology is not always agnostic to social or economic values and is capable of regulation by virtue of its architecture and design. Often referred to through the concept that 'code is law',[40] it highlights that value choices are made in the development of technology and its capabilities. Thus, technical value is high where the product or technology in question enables further innovation and behaves as a medium for further growth. This may be referred to as the 'generative' nature of the technology[41] that plays a role in its ability to become a general-purpose technology much like electricity, or the internet as we see it today.[42]

There are of course other concepts of value such as cultural, moral, political, geo-political, etc. However, the categorisation of social, economic, and technical subsumes the broadest range of issues and would suffice to deliberate on developing balanced regulations. While these categories are not water-tight and do influence each other, in policy discourse they present distinct objectives which are often lost when one gets prioritized over another. Some recent data governance regulations or proposals such as those restricting cross-border data flows,[43] emphasise the socio-political value of data and the geopolitical risks arising out of its misuse when located outside territorial jurisdictions. However, ICRIER's study on the economic implications of data flows found techno-economic reasons driving the choice of location for data storage.[44] The importance of data flows is amplified with the growth of data value chains spanning across organizations and countries. This has even prompted thinking along the lines of "global data value chains" and their implications.[45] The increase

39 MIT Technology Review Insights (n 29).

40 Lawrence Lessig, 'Code Is Law' [2000] *Harvard Magazine* <https://www.harvardmagazine.com/2000/01/code-is-law-html> accessed 13 November 2022.

41 Jonathan Zittrain, 'Law and TechnologyThe End of the Generative Internet' (2009) 52 Communications of the ACM 18 <https://dl.acm.org/doi/10.1145/1435417.1435426> accessed 14 November 2022.

42 George RG Clarke, Christine Zhenwei Qiang and Lixin Colin Xu, 'The Internet as a General-Purpose Technology: Firm-Level Evidence from around the World' (2015) 135 Economics Letters 24 <https://linkinghub.elsevier.com/retrieve/pii/S0165176515002773> accessed 14 November 2022.

43 Reserve Bank of India, 'Storage of Payment System Data - RBI/2017-18/153' <https://www.rbi.org.in/scripts/NotificationUser.aspx?Id=11244> accessed 14 November 2022.

44 Rajat Kathuria and others, 'Economic Implications of Cross-Border Data Flows' (Indian Council for Research on International Economic Relations 2019) <http://hdl.handle.net/11540/11375> accessed 13 November 2022.

45 Jeremmy Okonjo, 'Legal Constitution of Global Value Chains in the Digital Economy' (*Afronomicslaw.org*) <https://www.afronomicslaw.org/2020/11/11/legal-constitution-of-global-value-chains-in-the-digital-economy> accessed 12 November 2022.

in operational costs, compliance burden, the risk to privacy through potential surveillance of localised data etc. are consequences that are side-lined because of the prioritization of geo-political factors over techno-economic factors. On the other hand, where antitrust regulations for data centered businesses are more focused on economic value creation, social objectives of privacy violations and harm may become second-order priorities. An optimal data governance regime will promote collective value - an outcome of a country's economic, social, cultural, and political context. Necessarily, optimal data governance regimes will differ for countries.

Similarly, the selective assessment of values and binary choice can be seen in the debate between privacy and national security. For instance, in the context of encryption, tough adversarial views can result in an either-or type of policy options.[46] The friction between value perspectives was also visible through the ongoing Covid-19 pandemic. Contact tracing tools confronted stakeholders with a dilemma between the protection of individual privacy and protection against the pandemic[47] While several solutions emerged across the world, all of them dealt with tough trade-offs.

2. What should Data Governance in India look like?

Given that data governance has implications for diverse sections and sectors of society, and priorities that may range from economic growth, social development, and national security, it is not uncommon to see the presence of different ministries and departments[48] that try to regulate data with various competing objectives. Some of these regulations or policies are not specifically focused on data governance but impact the treatment of data significantly. For example, the e-commerce policy proposed by the Department for Promotion of Investment and Internal Trade while developing policy for the e-commerce sector had several proposals relating to data and its utility. It attempted to address the alleged unfair advantage that big-tech players had over small and upcoming players.[49] Proposals designed to facilitate

---

46 Yashovardhan Azad, 'Data Bill: The Security vs Privacy Debate' *Hindustan Times* (23 January 2021) <https://www.hindustantimes.com/analysis/data-bill-the-security-vs-privacy-debate-101611406252249.html> accessed 14 November 2022; Dipshikha Ghosh, 'The Big National Security versus Privacy Debate' *DNA India* (28 March 2017) <https://www.dnaindia.com/analysis/column-the-big-national-security-versus-privacy-debate-2371072> accessed 13 November 2022.

47 Bernard Marr, 'Why Contact Tracing Apps Will Be The Biggest Test Yet Of Data Privacy Versus Public Safety' *Forbes* <https://www.forbes.com/sites/bernardmarr/2020/06/01/why-contact-tracing-apps-will-be-the-biggest-test-yet-of-data-privacy-versus-public-safety/> accessed 13 November 2022; Ashkan Soltani Bergstrom Ryan Calo, and Carl, 'Contact-Tracing Apps Are Not a Solution to the COVID-19 Crisis' (*Brookings*, 27 April 2020) <https://www.brookings.edu/techstream/inaccurate-and-insecure-why-contact-tracing-apps-could-be-a-disaster/>; Pierfrancesco Lapolla and Regent Lee, 'Privacy versus Safety in Contact-Tracing Apps for Coronavirus Disease 2019' (2020) 6 DIGITAL HEALTH 205520762094167 <http://journals.sagepub.com/doi/10.1177/2055207620941673> accessed 14 November 2022.

48 For example, the Ministry of Electronics and Information Technology (MEITY) with the proposed data protection bill, the recently passed intermediary rules etc. also has numerous departments focusing on different areas, Ministry of Communications through its telecommunications regime, the Reserve Bank of India through its regulations for financial data, The Department of Consumer Affairs with proposed amendments for consumer protection on e-commerce platforms.

49 Ministry of Commerce and Industry, 'Draft National E-Commerce Policy: India's Data for India's Development' <https://dpiit.gov.in/sites/default/files/DraftNational_e-commerce_Policy_23February2019.pdf> accessed 13 November 2022.

different objectives of regulation are therefore laden with specific value judgments. The emergence of new regulatory instruments such as - mandating privacy by design, data minimization shows the prioritization of privacy rights over economic efficiency, or instruments such as data portability and interoperability, reflect the prioritization of market competition and economic opportunity. Many of these are still in proposal stages in India, while in other jurisdictions like Europe, are already in force through the GDPR and allied regulations.

India's digital journey and its existing heterogeneities, support the need for a comprehensive data governance regime that must accommodate varying and sometimes competing objectives. With an enormous population that is connected online issues such as privacy, security, consumer protection become important priorities. Yet, with an equally large population that is offline or unable to use the internet - issues of access, digital inclusion, and digital literacy become equally important demands on policy. The diversity of objectives can result in a disparate approach to policy making that will have sub-optimal outcomes. Similarly, divergent positions of stakeholders or stakeholder groups such as the interests of small business as against large establishments, foreign and local businesses, private and public enterprises also complicate policy formulation. While several parts of the government machinery have responded to these diverse needs, the related regulations developed in India are neither coordinated nor comprehensive. In the last year alone, India has seen attempts by various sectoral regulators or different ministries that have introduced regulations either with a very narrow view such as RBI's new guidelines for recurring payments through credit cards or broad-based regulations that may be counterproductive.[50] For example, e-commerce rules under the Consumer Protection Act that tried to accommodate both competition concerns and consumer protection objectives, instead created potential contradictions with other regulations.[51] While the eagerness to address problems is commendable, the lack of patience seen through ad-hoc unpredictable policy development processes and a lack of a larger vision to plan and implement strategically is disappointing. A simple case in point being the recently implemented intermediary guidelines[52] which has created a chimera of the Ministry of Information and Broadcasting and the Ministry of Electronics and Information Technology for regulating online

---

50 Ashwin Manikandan and Saloni Shukla, 'New Norms Put Auto Debits in No Man's Land' *The Economic Times* (21 October 2021) <https://economictimes.indiatimes.com/tech/technology/new-norms-put-auto-debits-in-no-mans-land/articleshow/87169930.cms> accessed 14 November 2022.

51 As a platform, e-commerce entities would come under the ambit of 'intermediary' as envisaged under the Intermediary Guidelines under the IT Act and imposing personal liability would conflict with safe harbour provisions under section 79. On issues of *mis-selling and mis-leading advertisements*, the proposed amendments tend to conflict with provisions of the parent legislation i.e., the Consumer Protection Act, 2019. Specifically, the parent legislation under s.21 (6) provides some defence for potentially mis-leading advertisements that may have been part of '*ordinary course of business*'. Similarly, the inclusion of liability for innocent misrepresentation through the proposed amendments are not aligned with the objectives of protecting consumers against malicious and fraudulent misrepresentations by the parent act.

52 Ministry of Electronics and Information Technology, 'Notification Dated, the 25th February, 2021 G.S.R. 139(E): The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021' <https://www.meity.gov.in/content/notification-dated-25th-february-2021-gsr-139e-information-technology-intermediary>.

content. The guidelines attempt to regulate not only social media intermediaries but also digital news media in the same stroke.[53] This trend of knee-jerk regulations implemented so far is not only complex and overlaps in the jurisdiction but can result in competing or conflicting objectives and outcomes. The most confounding aspect in this quagmire is that all these developments occur while the country awaits a primary data protection legislation. The proposed personal data protection bill has been deliberated for years, languished between parliamentary committees, and tabled in the parliament, only to be withdrawn by the government. At the time of writing this essay, the Indian government has stated that it's reworking the data protection bill along with a slew of legal reforms for *"contemporary and future challenges and catalyse Prime Minster Narendra Modi's vision of India Techade."*[54] While India has been very active in its regulatory responses, the fundamental approach is still unclear.

For better inspiration, we may look to Singapore that showcased an excellent example of balance of values, nuanced regulatory clarity and cooperation between regulatory agents and other stakeholders. A well-synthesised and potentially optimal data governance regime is visible in Singapore's proposed data portability framework. A significant amendment in 2020 to the Singapore Personal Data Protection Act introduced a data portability requirement among others in its first set of significant changes to its legislative framework since 2012.[55] This was jointly developed by the Personal Data Protection Commission and the Competition and Consumer Commission of Singapore. This collaboration is a reflection of a comprehensive approach that has laid down economic, social and technical requirements in the design of the regime. The discussion paper issued provides a thorough examination of different dimensions of introducing a data portability requirement ranging from competition implications, and data protection concerns. It captured the potential costs involved in compliance, and potential barriers to entry that such provisions may create. The discussion paper was followed by public consultation on the introduction of data portability requirement and other amendments to the PDPA of Singapore.[56] Following the nearly two-month comment period that closed in July

---

53 Aashish Aryan, 'Explained: Social Media and Safe Harbour' *The Indian Express* (27 May 2021) <https://indianexpress.com/article/explained/intermediary-guidelines-digital-media-ethics-code-facebook-twitter-instagram-7331820/> accessed 13 November 2022; Malavika Raghavan, 'India's New Intermediary & Digital Media Rules: Expanding the Boundaries of Executive Power in Digital Regulation' (*https://fpf.org/*) <https://fpf.org/blog/indias-new-intermediary-digital-media-rules-expanding-the-boundaries-of-executive-power-in-digital-regulation/> accessed 14 November 2022; Internet Freedom Foundation, 'Explainer: Why India's New Rules for Social Media, News Sites Are Anti-Democratic, Unconstitutional' *Scroll.in* (25 February 2021) <https://scroll.in/article/988105/explainer-how-indias-new-digital-media-rules-are-anti-democratic-and-unconstitutional> accessed 13 November 2022.

54 BS Reporter & PTI, 'Govt Withdraws Data Protection Bill, 2021, Will Present New Legislation' *Business Standard* (3 August 2022) <https://www.business-standard.com/article/economy-policy/centre-withdraws-personal-data-protection-bill-2019-to-present-new-bill-122080301226_1.html> accessed 13 November 2022.

55 'PDPC | Amendments to the Personal Data Protection Act and Spam Control Act Passed' <https://www.pdpc.gov.sg/News-and-Events/Announcements/2020/11/Amendments-to-the-Personal-Data-Protection-Act-and-Spam-Control-Act-Passed> accessed 14 November 2022.

56 Personal Data Protection Commission, Singapore, 'Public Consultation on Review of the Personal Data Protection Act 2012 - Proposed Data Portability and Data Innovation Provisions' (2019) <https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Legislation-and-Guidelines/PDPC-Public-Consultation-Paper-on-Data-Portability-and-Data-Innovation-Provisions-(220519).pdf?la=en> accessed 14 November 2022.

2019, the PDPC published all responses received during the public consultation[57] and in January 2020 provided a detailed document which curated inputs and the point-by-point responses of the PDPC.[58] It took into account feedback received from stakeholders and communicated the intent to reduce the scope of data that would come under the scope of the portability provision while retaining the requirement itself. Furthermore, the amendments also excluded the exemptions previously available to private actors acting on behalf of government. A notable shift is also seen in the approach to consent – with greater clarity on consent requirements, new exceptions, and development of the concept of deemed consent. Furthermore, the process factored in economic considerations involved in facilitating data portability including different approaches for valuing data.[59] What we wish to highlight here is not specific provisions to mimic but a reliable process that assess diverse value conceptions and inspires balanced regulations. A predictable and transparent process also contributes to a sense of accountability of actors that are dealing with citizens' personal information.

Some may point to India's recently initiated Account Aggregator (AA) Network as a good example of balanced value realisation. The framework is considered to have managed to create a financial data sharing entity that facilitate seamless sharing of data for a consumer to avail financial services based on a strong consent requirement.[60] The AA as an entity would not be able to create user profiles nor have access to the contents of the data they transfer and would be under RBI's strict regulations. Yet, there are strong criticisms that identify several concerns relating to ethics, consent friction and fatigue, lack of specificity in associated guidelines to prevent abuse, storage of data after revocation of consent etc.,[61] clearly highlighting that the choice of solutions has prioritized one set of values over another or failed to meaningfully address specific concerns.

---

57 ibid.

58 Personal Data Protection Commission, Singapore, 'Response to Feedback on the Public Consultation on Proposed Data Portability and Data Innovation Provisions' (2020) <https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Legislation-and-Guidelines/Response-to-Feedback-for-3rd-Public-Consultation-on-Data-Portability-Innovation-200120.pdf?la=en> accessed 14 November 2022.

59 INFOCOMM Media Development Authority, 'Guide to Data Valuation for Data Sharing' <https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Guide-to-Data-Valuation-for-Data-Sharing.pdf> accessed 14 November 2022.

60 George Matthew, 'Account Aggregators: New Framework to Access, Share Financial Data' (*The Indian Express*, 6 September 2021) <https://indianexpress.com/article/explained/account-aggregators-new-framework-to-access-share-financial-data-7490966/> accessed 14 November 2022; 'Account Aggregator, a Game Changer' (9 September 2021) <https://www.thehindubusinessline.com/opinion/account-aggregator-a-game-changer/article36385799.ece> accessed 14 November 2022; ET Bureau, 'A New Digital Revolution in India' *The Economic Times* (8 September 2021) <https://economictimes.indiatimes.com/opinion/et-editorial/a-new-digital-revolution-in-india/articleshow/85955574.cms> accessed 14 November 2022; Rajat Deshpande, 'How Data Can Change Lending Experience in India?' (*cnbctv18.com*, 22 December 2021) <https://www.cnbctv18.com/finance/how-data-can-change-lending-experience-in-india-11902902.htm> accessed 14 November 2022.

61 Rohan Jahagirdar and Praneeth Bodduluri, 'Digital Economy: India's Account Aggregator System Is Plagued by Privacy and Safety Issues' (2020) 55 Economic and Political Weekly <https://www.epw.in/engage/article/digital-economy-indias-account-aggregator-system> accessed 26 October 2022.

These instances highlight how issues of regulation in the digital economy are no longer disjoint or completely separable. The overlap of jurisdictions has become inevitable. There is an evident need for integration for law making and regulatory responses. This is no easy task. A converged regulator brings many benefits but also poses questions of effective enforcement and efficiency. This becomes especially challenging in the context of India's institutional history that is rife with turf battles. Given a regulatory landscape that is dotted with several key stakeholders and their varying objectives, a multistakeholder approach to governance becomes necessary. A statutory warning that come with the support for this approach is that it is deeply detrimental in the absence of specific measures for inclusion of stakeholders, transparent and participative processes, and accountability of actors. Without these factors ensured, any multistakeholder approach would merely become a conduit for the will of the loudest or the most powerful voice in the ecosystem.

India's approach to multistakeholder participation in policymaking is mixed. Some institutions like the Telecom Regulatory Authority of India (TRAI) are by design consultative. Yet the lack of an institutionalized and predictable process of multistakeholder policymaking is clearly felt. The policy journey in the development of the data protection legislation for example began with the Justice Srikrishna Committee's consultative processes, however, upon moving to the stage of the bill being reviewed by a Joint Parliamentary Committee, it became invite-only with closed-door consultations. While there are arguments both for and against open or closed consultative processes, the trouble is a lack of consistency and predictability of the process in addition to the lack of accountable institutions that facilitate policy development. Further, this problem was only amplified by the subsequent withdrawal of the bill. In the absence of consistent, predictable, and transparent processes, stakeholders are left in a cliff-hanger of suspense relying on media reports quoting unnamed sources[62] and statements from ministers[63] to guess both the unpredictable timelines and undecided scope of the future of data protection laws in India. In this context, it is important to be reminded that the benefits of an institutionalized, consistent, and predictable process include better transparency and accountability of stakeholders in the process. This can also improve trust not only among stakeholders involved in the regulatory ecosystem but also among general citizens and motivate wider civic engagement. It would facilitate a democratic and reliable process to identify and deliberate value(s) that need to be considered holistically in making policy choices. Eventually, the goal is not a utopian realization of all values or achieving an average in a comprehensive regulation. It is to provide the opportunity

62 Aishwarya Paliwal, 'New Data Protection Bill to Be Stringent, Tech Giants in India Need to Strictly Follow Rules: Govt' *India Today* <https://www.indiatoday.in/technology/features/story/new-data-protection-bill-to-be-stringent-tech-giants-in-india-need-to-strictly-follow-rules-govt-1986127-2022-08-10> accessed 14 November 2022.

63 PTI, 'Significant Work Done, Draft Digital India Act Framework by Early 2023: MoS IT' *The Hindu* (6 November 2022) <https://www.thehindu.com/business/Economy/significant-work-done-draft-digital-india-act-framework-by-early-2023-mos-it/article66103357.ece> accessed 14 November 2022.

for adequate representation and consideration of stakeholder interests through transparent deliberations. This would facilitate a review of different permutations and combinations of value frameworks and enable informed decision-making in a nuanced manner especially when it comes to trade-offs. Furthermore, developing regulations for the digital ecosystem has significant repercussions globally - for people, businesses, and the technology itself. Given India's quest for leadership in global affairs, it would need to improve its policymaking processes and institutions to factor not only national perspectives but build a collaborative approach with different jurisdictions and their value frameworks.

Conclusion

We reiterate the undercurrent of this essay that echoes several other scholars and continues to be attempted by various stakeholders - move away from the binaries in assessing the value of data while developing the data governance ecosystem. In India's pursuit of a balanced data governance regime so far, the dominant focus has been on the specific substance of regulation that often fail to accommodate diverse value drivers. Whereas the need of the hour is to invest in better institutions and reliable processes that can holistically approach the value(s) assessment of data. Championing one set of values while disregarding another is a recipe for unsustainable ecosystems that breed inequality and inefficiency. We are at an inflection point in history that asks several questions and raises the stakes so high that decisions today will determine the kind of future we build in a world where lines between digital and analog fade.

Data Stewardship: Re-imagining Data Governance

Astha Kapoor[1]

Introduction

In 2018, user data of millions of Facebook users was collected without consent and leaked to Cambridge Analytics for political advertising.[2] Incidents like this, and many more over the last few years, demonstrate how personal data can be leaked, harvested and used for micro-targeting and behavioural manipulation. Individuals, despite the growing number of data protection regulations are unable to negotiate or find redressal in any meaningful way thus beckoning the need for an alternative approach to data governance. The new approach should empower individuals to participate more meaningfully in the conversation on how their data is collected, used and shared. This paper proposes a new approach to data stewardship, that may help address some of the major questions of data governance on the limitations of the current focus on protection, consent. Simply put, a data steward is an independent intermediary who acts on behalf of those whose data it is and those that are affected by the use of that data and helps unlock the value of data while safeguarding the rights of generators. Data stewards can be delegated consent, and are duty-bound to act in the interest of those they represent. There are multiple ways in which data stewards enact this duty – through fiduciary responsibility of care and loyalty, to make sure that data stewards do not exploit the data in question, and always act in the best interest of communities.[3]

The insufficiency of existing data protection frameworks is clear – whether the GDPR in the European Union (EU) or the now withdrawn,[4] Data Protection Bill, 2021 (DPB 2021) in India, are focused on privacy, and engage citizens only through notice and consent i.e. individuals are notified about their data collection, use of data collection is declared and people can choose to consent or not. However, these notices are complex and hard to navigate and as a result the quality of consent is poor i.e. individuals are not able to provide informed consent and tend to agree to

---

1 Astha is the Co-Founder of Aapti Institute, Bangalore. She can be reached at astha@aapti.in.

2 Nicholas Confessore, 'Cambridge Analytica and Facebook: The Scandal and the Fallout So Far' *The New York Times* (4 April 2018) <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html> accessed 4 November 2022.

3 Siddharth Manohar, Aditi Ramesh, and Astha Kapoor, 'Understanding Data Stewardship: Taxonomy And Use Cases' (Aapti Institute 2020) <https://thedataeconomylab.com/2020/06/24/data-stewardship-a-taxonomy/> accessed 4 November 2022.

4 Soumyarendra Barik, 'Explained: Why the Govt Has Withdrawn the Personal Data Protection Bill, and What Happens Now' *The Indian Express* (4 August 2022) <https://indianexpress.com/article/explained/explained-sci-tech/personal-data-protection-bill-withdrawal-reason-impact-explained-8070495/> accessed 7 November 2022.

problematic data sharing arrangements.[5] Further, existing frameworks are focused on individual rights and do not consider data as a collective, communal experience – the data about individuals impacts groups, and data about groups impacts individuals – data is a relational good, and therefore the rights around it should be viewed through that lens which current frameworks do not do. There is one exception to individualised thinking, the Non-Personal Data Committee Report (NPD Report)[6] in India which discusses community rights to data and aims to empower communities to both extract value from and prevent harms that come from data. However, the conversation on community data rights as proposed in the NPD Report has been derailed with the last data protection bill (DPB 2021) recommending that personal and non-personal data should be regulated by the same Data Protection Authority – given that the bill has been withdrawn, there may be opportunity to rephrase this conversation.[7] The last version of the bill moved the conversation away from value of data, and focuses on protection;[8] While the shift to protection is welcome and important, the mechanisms for enhancing community rights, and distributing value more evenly to collectives also vanishes.

These protection based frameworks do not fundamentally challenge the core issues of the data economy - a mismatch in power between individuals/communities and platforms, where data is extracted for the benefit of technology companies in a way where people suffer harms due to loss of privacy.[9]

The focus on privacy and protection also prevents individuals from sharing their data for purposes that might benefit public causes and create a broader social good. Individuals don't have the ability to direct data towards issues they may care about, or want to surface through research and innovation. Even civil society organisations are unable to access data of individuals and communities to help deliver on social causes, as data is not made available to the generators. For instance, patients of multiple sclerosis may want to collectivise their data and use it for specific questions they want answered but do not have avenues to do so. Therefore, there is a need for new data governance mechanisms that are structured around empowerment, agential rights and collective bargaining are required.

---

5 Claire Park, 'How "Notice and Consent" Fails to Protect Our Privacy' (*New America*, 23 March 2020) <http://newamerica.org/oti/blog/how-notice-and-consent-fails-to-protect-our-privacy/> accessed 4 November 2022.

6 Committee of Experts, 'Report by the Committee of Experts on Non-Personal Data Governance Framework' (The Ministry of Electronics & Information Technology 2020) <https://ourgovdotin.files.wordpress.com/2020/12/revised-report-kris-gopalakrishnan-committee-report-on-non-personal-data-governance-framework.pdf> accessed 4 November 2022.

7 Apoorva Mandhani, 'Non-Personal Data, Social Media — What New "data Protection Bill" Could Look Like' *ThePrint* (6 December 2021) <https://theprint.in/theprint-essential/non-personal-data-social-media-what-new-data-protection-bill-could-look-like/776389/> accessed 4 November 2022.

8 Internet Freedom Foundation, 'Key Takeaways: The JPC Report and the Data Protection Bill, 2021 #SaveOurPrivacy' (Internet Freedom Foundation 2021) <https://internetfreedom.in/key-takeaways-the-jpc-report-and-the-data-protection-bill-2021-saveourprivacy-2/> accessed 4 November 2022.

9 Global Partnership on Artifical Intelligence, 'Enabling Data Sharing for Social Benefit through Data Trusts - GPAI' <https://gpai.ai/projects/data-governance/data-trusts/> accessed 7 November 2022.

In this context, data stewardship could help address both the issues of participation and enhanced decision-making powers for individuals and communities, and enable people to better use their data for public value.

Data stewardship is being explored in different shapes and forms in conceptual conversations in policy documents such as the NPD Report in India to on-ground experiments that are building models to help actualise these different models of data stewardship across the globe, and the functions they might perform. At the outset, the paper outlines the utility and limitations of some of the most popular/ common models of data stewardship in the current digital ecosystem. It also discusses data trusts, as mentioned in the NPD Report, and evaluates its effectiveness in the Indian landscape. Finally, the paper will conclude with the challenges of establishing mechanisms of stewardship in India and outline some of the key requirements that can help enable stewardship over the next few years.

1. Some Models of Data Stewardship

1.1. Data Cooperatives

One of the most common models of data stewardship is "data cooperatives". Cooperatives, which have a rich global history, are structures where members make collective decisions on shared assets. Cooperatives exist for housing, and cooperatively run companies and hospitals which are very effective. Cooperatives are usually voluntary, are membership driven and decisions are made democratically – one member, one vote, and all economic surplus is shared equitably among members. The cooperative model is being increasingly applied to data since the model allows individuals to pool data, and co-govern it to exercise greater rights, seek collective representation and generate value for the community, and public more widely. It is also attractive for organisations acquiring data, as it is easier to interface with a collective, rather than individuals.[10] Further, data cooperatives are effective for collective bargaining, directing data value and enhancing broader resistances to the ways in which data is collected and used. A much cited example of this model is Driver's Seat,[11] a San Francisco based cooperative that gives its members, Uber and Lyft drivers, visibility into how their mobility data is being used by platforms they work for. Driver's Seat has an app through which workers can submit their location, working hours and earning information and receive insights that can help them understand their potential earnings and performance and enable them to make more informed choices on which platforms can enhance their income.

10 Sameer Mehta, Milind Dawande, and Vijay Mookerjee, 'Can Data Cooperatives Sustain Themselves?' (*LSE Business Review*, 2 August 2021) <https://blogs.lse.ac.uk/businessreview/2021/08/02/can-data-cooperatives-sustain-themselves/> accessed 4 November 2022.

11 'Driver's Seat Cooperative' (*Driver's Seat*) <https://driversseat.co/> accessed 4 November 2022.

Driver's Seat aggregates data from its members and sells it to municipalities, and distributes the earnings to members. The app enables an understanding of the relational value of data, and collectivises driver experiences which can be over time, used to negotiate for better worker rights with platforms. There are other examples, such as Open Data Manchester, that is building an energy data cooperative to help consumers engage more meaningfully with providers, and manage usage and payments better.[12]

While India has a rich history of cooperatives in different sectors, the idea of data cooperatives has not found traction at the moment. This is because organising and mobilising around questions of data is nascent and complex for multiple reasons – first, the harms from and value of data is poorly understood by people; second, shared experiences with data and relatedly, the expectations from what decisions about data should yield are different within communities, third, the technological and human capacity required to set up data cooperatives is limited and finally, the business model for data cooperatives is an open question which needs to be solved for. However, there is a case to be made for existing cooperatives to think laterally about data related functions, as data usage becomes more and more ubiquitous. For instance, farmer credit cooperatives can consider collectively governing certain kinds of data that impact productivity and in turn reflect the types of loans cooperatives are able to receive to ensure tailored financial products that do not rely on data extraction – data cooperatives can benefit from the homogeneity and strong ties shared by members of cooperatives. But, to make data cooperatives a reality in India, there is a need to educate, train and inform institutions on the value of collective data governance, and the role of cooperatives in enabling this.[13] In Europe, Worker Info Exchange, enable gig workers to receive a copy of the data generated about them by platforms and submit "Subject Access and Data Portability Requests"[14] to support workers in challenging unfair decisions, and are in the process of setting up a data trust.[15] The recommendation for data portability in the report presented by the Joint Parliamentary Committee (JPC)[16] allows data principals to port their data (access and transfer their data

12 'Home' (*Open Data Manchester*) <https://www.opendatamanchester.org.uk/> accessed 4 November 2022.

13 Julian Trait, Open Data Manchester, 'The Case for Data Cooperatives' (Aapti Institute 2021) <https://thedataeconomylab.com/2021/09/06/the-case-for-data-cooperatives/> accessed 4 November 2022.

14 'Request Data' (*Worker Info Exchange*) <https://www.workerinfoexchange.org/request-data> accessed 4 November 2022.

15 Worker Info Exchange, 'Gig Workers Score Historic Digital Rights Victory against Uber & Ola' (*Worker Info Exchange*, 15 March 2021) <https://www.workerinfoexchange.org/post/gig-workers-score-historic-digital-rights-victory-against-uber-ola-2> accessed 7 November 2022.

16 Pragni Kapadia, 'New Data Protection Regime in the Making in India' *Financial Express* (8 January 2022) <https://www.financialexpress.com/opinion/new-data-protection-regime-in-the-making-in-india/2401518/> accessed 7 November 2022.

from companies that hold it)[17], an important tool in preventing harms against big tech.[18] This right to port, as demonstrated by Worker Info Exchange can be used to port data to data cooperatives.

1.2. Data Commons

Data commons are a mechanism to pool data and govern it as a common resource. The idea of commons as applied to data comes from Elinor Ostrom's work on public goods which points to flexible structures that allow communities to manage resources in accordance with their own unique rules and values.[19] Governance of commons in technology is widely seen in efforts such as Wikipedia, Open Street Maps where open access objects are managed by a broader community invested in its upkeep.

Data Commons and Data Cooperatives are similar to the point that shared resources are co-governed, but membership and governance mechanisms are much more informal in the former.[20] Commons may be governed through different systems that communities may evolve together - these could take the form of a cooperative or foundation, or other institutional mechanisms. Communities may also evolve systems for collective governance, dispute resolution and redressal in case of harm. Data Commons represent a fluidity of decision-making and initiative on behalf of the community, and not the application of an "off the shelf" model of data stewardship. They may be seen as a starting point to launch into defined models like cooperatives, trusts etc.

Governing through a commons approach is implemented in India in the management of natural resources, for instance water. There is evidence to suggest that establishing water bodies as common resources and strengthening community stewardship results in enhancing water conservations and managing demand.[21] For India, data commons can be a starting point to think through more formal and familiar structures such as data cooperatives, which can be adapted to data once as commons way of dealing with data is understood. To imagine data as a shared resource requires significant sensitisation both about

---

17 Bennett Cyphers and Cory Doctorow, 'Privacy Without Monopoly: Data Protection and Interoperability' (Electronic Frontier Foundation 2021) <https://www.eff.org/document/privacy-without-monopoly-data-protection-and-interoperability> accessed 4 November 2022.

18 Internet Freedom Foundation, 'Key Takeaways: The JPC Report and the Data Protection Bill, 2021 #SaveOurPrivacy' (Internet Freedom Foundation 2021) <https://internetfreedom.in/key-takeaways-the-jpc-report-and-the-data-protection-bill-2021-saveourprivacy-2/> accessed 4 November 2022.

19 Anouk Ruhaak, 'Data Commons & Data Trust' (*Medium*, 15 May 2020) <https://medium.com/@anoukruhaak/data-commons-data-trust-63ac64c1c0c2> accessed 4 November 2022.

20 ibid.

21 Foundation for Ecological Security, 'Water Commons - Influencing Practice and Policy' (Foundation for Ecological Security 2017) <https://fes.org.in/resources/impact/internal-reports/water-commons-influencing-practice-and-Policy-social-return-on-investment-report-2016-17.pdf>.

the collective harms and community value of data. There is much to be learnt from indigenous communities in Canada and Australia that are pursuing data sovereignty and governing through commons principles to reclaim data and direct it to benefit the community.[22]

1.3. Personal Data Stores (PDS)

PDS or data vaults are stewards that are used to securely store user data, and give the individual the right to decide who can access their data. PDS can offer multiple services but most commonly serve as a way to manage consent, and aggregate personal data from different sources (social media, mobility apps, fitness apps etc) and make it visible and useful to individuals. Some PDS also offer individuals the opportunity to monetise their data.[23] There are multiple start-ups building PDS companies, enabling individuals to exercise greater control over their data. An example of a PDS is Digi.me,[24] an app which allows users to upload data and connect social media apps to store their data and decide which data points are shared with which parties. Digi.me provides visibility on data access and final decision making to the users. Data is encrypted and stored at a platform of the consumer's choice and is only accessible once consent is given. Tim Berner Lee's Solid also enabled "data ownership", enabling users to decide where to store and whom to share their data with. Solid relies on decentralised storage systems that prevent vendor lock-in and enable switching between providers.[25]

India's DigiLocker allows individuals to store their documents (personal data) but does not provide any advisory services that are critical for the role of a steward. Further, DigiLocker has demonstrated certain vulnerabilities such as its now resolved authentication flaw, which make it potentially unreliable for users.[26] Similarly, consent managers as mentioned in the Data Empowerment and Protection Architecture (DEPA) allow individuals to manage their consent – to both give and revoke access to financial data, which is a role a personal data store also performs and enable transparency on what data is being collected by whom. However, consent managers are attractive because they are data blind,[27] and a passthrough for data moving from point A to point B. However, their design at

22 Stephanie Russo Carroll and others, 'The CARE Principles for Indigenous Data Governance' (2020) 19 Data Science Journal 43 <http://datascience.codata.org/articles/10.5334/dsj-2020-043/> accessed 4 November 2022.

23 Siddharth Manohar, Aditi Ramesh, and Astha Kapoor (n 3).

24 'The Digi.Me Platform' <https://digi.me> accessed 4 November 2022.

25 'Solid' <https://solid.mit.edu/> accessed 4 November 2022.

26 Jagmeet Singh, 'Flaw in DigiLocker Put Over 3.8 Crore Accounts at Risk: Researcher' *Gadgets 360* <https://www.gadgets360.com/internet/news/digilocker-vulnerability-3-8-crore-accounts-hack-documented-2239419> accessed 4 November 2022.

27 Ratul Roshan and Aparajita Srivastava, 'Consent Managers in the Financial Space: Account Aggregators | Ikigai Law' (11 November 2020) <https://www.ikigailaw.com/consent-managers-in-the-financial-space-account-aggregators/> accessed 4 November 2022.

the moment do not serve to provide individuals information on what data should be shared or not, and are currently limited only to the financial services use case. That said, consent managers have immense potential to evolve into personal data stewards for individuals, working with them to manage data access and use, across multiple use cases such as health, employment along with finance.

1.4. Data trusts

Data trusts are the most talked about and least understood mechanism of stewardship. Trusts are designed to have a trustee of data rights, who has fiduciary responsibility towards a group of beneficiaries. Trustees have a duty of care and loyalty towards beneficiaries and are empowered to negotiate on behalf of the collective.[28] Data trusts must always have a clear purpose, that is communicated to all beneficiaries such as holding data in trust to provide cyclists in Bengaluru better routes.

Trusts law, which data trusts draw from, exists in some parts of the world – UK, US and Canada. In civil law jurisdictions such as in Germany, fiduciary responsibility needs to be codified through contracts.[29] Data trusts, by definition, allow for collective stewardship through delegation of decisions to the trustees, and differ from Cooperatives, where decisions are taken collectively by the community. In summary, functionally, data trusts must perform three key functions, enable data driven innovation for social and economic benefit, rebalance power asymmetries in data exchange, and anticipate, prevent and manage vulnerabilities from data use.[30]

Data trusts are getting a lot of traction as they are a formal structure that provides accountability, a way to pool data rights and a platform to collectively negotiate questions of harm and value. However, real world evidence on formally institutionalised data trusts is limited. The concept has been misappropriated on multiple occasions, most notably by the now abandoned Sidewalk Labs project which aimed to establish a data trust without the requisite accountability and participation mechanisms that are typical of the model. Sidewalk's Urban Data Trust lacked clarity on how the data trust would be used to serve the community, and didn't necessarily safeguard citizen interests of data protection.[31] That said,

28 Aapti Institute and Open Data Institute, 'Enabling Data Sharing for Social Benefit through Data Trusts' <https://docs.google.com/document/d/18HPZbsd9DLQp5fk7iSzS6fs-ptGiWSJrm34UdR_3aMg/edit#heading=h.3q3bsupsw620> accessed 4 November 2022.

29 Aapti Institute, 'Aapti: Enabling Data Trusts - Output 2 (Legal Review)' <https://docs.google.com/document/d/1xdQsPNxRRdxmouzgPpkFIMB-phKBIbEnMS4SKaQTBMk/edit?usp=sharing&usp=embed_facebook> accessed 7 November 2022.

30 Neil Lawrence, Jessica Montgomery, and Seongtak Oh, 'Enabling Data Sharing for Social Benefit through Data Trusts at the Global Partnership on AI, GPAI' (*OECD.AI - Policy Observatory*, 3 August 2021) <https://oecd.ai/en/wonk/data-sharing-data-trusts> accessed 4 November 2022.

31 Teresa Scassa, 'Designing Data Governance for Data Sharing' [2020] Technology and Regulation 44 <https://techreg.org/article/view/10994> accessed 4 November 2022.

there are a few enablers critical to actualise data trusts in any jurisdiction – data protection frameworks, data sharing frameworks and some semblance of fiduciary obligations.[32]

The NPD Report in India is one of the first policy documents to enunciate the importance of collective rights over community Non-Personal Data (NPD) – this is in contrast with the dominant individual led protection frameworks. The NPD Report[33] recognises beneficial interests over community data. It identifies five key principles to ascertain community rights over data: (i) a community's right over resources associated collectively with it; (ii) consent of the community for use of such resources; (iii) benefit sharing with the community; (iv) transparency in recording community resources to prevent misuse and enable easy access of the legitimate kind; and (v) community's participation in governance of community resources. The NPD Report also recommends the creation of 'data trustees' as intermediaries to exercise rights on behalf of the group/community. The committee sources this community right from Article 39(b) and (c) of the Indian Constitution (Directive Principles of State Policy) which stipulates that the ownership and control of resources ought to be distributed to serve the common good and to prevent the concentration of wealth. However, the roles and functions of data trustees are not clearly specified, and are not in tandem with the broader understanding of data trusts, for instance the fiduciary responsibilities of data trustees is unclear.[34]

More functionally, the NPD Report, while specifying that non-profits can be data trustees, does little to explain how these structures would be established and work with different communities on the ground. Establishing stewardship is complex, and needs to happen at different levels. To make stewardship a reality, three key components are required – a robust data protection framework, avenues for secure data sharing and a cogent articulation of fiduciary responsibilities that can be applied to data stewards.[35] Despite the strong push by the NPD Report to establish data trustees, unfortunately, the enabling architecture does not exist. India fares poorly on data protection, and has not passed a data protection framework without which data stewardship cannot be actualised. That said, the DPB 2021 clubbed NPD with personal data protection, and made no mention of the use of data trusts. The conversation on data trusts and their use seems to have been put on the back burner for now, as once the Bill is passed into law,

32 Aapti Institute (n 29).

33 Committee of Experts (n 6).

34 Aapti Institute, 'Comment on the Revised Report by the Committee of Experts on Non-Personal Data Governance Framework' <https://thedataeconomylab.com/wp-content/uploads/2021/02/Comment-on-the-Revised-Report-on-NPD-Governance-Aapti-Institute.docx.pdf> accessed 4 November 2022.

35 Neil Lawrence and Seongtak Oh, 'Enabling Data Sharing For Social Benefit Through Data Trusts' (Aapti Institute and Open Data Institute 2021) <https://gpai.ai/projects/data-governance/data-trusts/enabling-data-sharing-for-social-benefit-through-data-trusts.pdf> accessed 4 November 2022.

the focus is likely to be basic protection frameworks, and not enabling models of stewardship that will need the intellectual and legal scaffolding the DPB 2021 provided.

Currently, data protection is regulated by the Information Technology Act, 2000 (IT Act), and the rules under it. While the IT Act's conception of data protection is quite limited in scope, only recognising consent, access, and correction rights, the DPB 2021, drafted in the backdrop of the landmark judgement that recognised the right to privacy, extended additional rights like the right to erasure and portability. The DPB 2021 also envisaged giving data principals the right to delegate the exercise of their agency (provide or withdraw consent) to a new category of data fiduciaries termed as consent managers.[36] Collective rights without individual rights are not possible, and therefore the NPD's recommendations are moot.

While India has codified trusts and trustee's fiduciary responsibilities and how trustees are selected, the feasibility for legal trusts to hold data rights as the subject matter lacks legal certainty. Moreover, in the absence of dedicated data protection legislation, India's recognition of individual rights over personal data remains weak, further restricting the possibility of data trusts' to act as intermediaries.

2. One Size Fits All?

This list of models isn't exhaustive, other mechanisms such as Data Collaboratives, Exchanges[37] exist and offer different degrees of accountability and participation to people. There are also data advocates that actively work with communities to sensitise them on the value of data stewardship and help communities build models of stewardship. Irrespective of the model, data stewards are critical instruments for rebalancing power in the data economy, and making data available for research, advocating for data rights, enabling collective bargaining on data rights and ensuring transparency on decisions. Given the diversity of functions data stewards perform, it is complex to recommend one model – the choice of model is determined by the community, purpose it aims to fulfil, data type it stewards among other variables determine the form and function of the steward.

The description of different models demonstrates that purpose is important to understand, as is the will and need of the community. For instance, communities with shared experiences and the ability to collectively decide may opt for data

36 Report of the Joint Committee on the Personal Data Protection Bill, 2019 (16 December 2021), <http://loksabhaph.nic.in/Committee/CommitteeInformation.aspx?comm_code=73&tab=1>.

37 Siddharth Manohar, Aditi Ramesh, and Astha Kapoor (n 3).

cooperatives, whereas communities may feel more comfortable delegating more complex decisions on data use to a trusted intermediary in the form of a data trust, that is legally bound to decide in the best interest of the whole community and can be held accountable in case of misuse or harm.

This also links to the level of participation the community would like to engage in – some communities may choose to remain informed (through existing mechanisms of notice and consent) while others may want to be consulted (through data cooperatives) and still others may want to actively engage to draw empowerment from the steward (through data trusts and cooperatives). Data stewardship models must provide the flexibility and choice for communities to pick and choose their form of engagement, and these decisions may need to be dynamic so that people have the option of changing their participation style, if required. Fundamentally, the community must have a choice on the degree and type of participation that is most comfortable and suitable.

Similarly, the purpose of the steward is critical in determining the form it takes. If the objective is to unlock the mobility data of a collective, and provide oversight – then a loosely defined model like data commons is likely to be most appropriate, as levels of accountability, participation are both low and flexible. However, if the objective of the data steward is to ensure that worker data is not being collected once they've logged off the application, then a more formalized steward like a cooperative, or trust may be required to ensure that the objectives are fulfilled.

Other factors such as data type (personal, non-personal) also determine the form of a steward as does the sector (mobility data is different from health data). The fundamental lesson from analysing different models of data stewardship is that *there is no one-size fits all,*[38] every community has to evolve its own model of stewardship based on its own needs. The flexibility of this approach will ensure that the landscape remains need based, bottom up and community centric.

There are also questions about the business model of data stewards, and how to ensure that these trusted intermediaries remain focused on the welfare of the people they represent, and do not fall trap to perverse business models anchored in data extraction.[39] For instance, if data cooperatives rely on member fees, the founder of Good Data Cooperative (now defunct) believed that it would take 500,000 members for the organisation to be sustainable.[40] These are large numbers and require a scale data stewardship models have not achieved so far – as ideas remain small and

38 Sylvie Delacroix and Neil Lawrence, 'Disturbing the "One Size Fits All", Feudal Approach to Data Governance: Bottom-Up Data Trusts' [2018] SSRN Electronic Journal <https://www.ssrn.com/abstract=3265315> accessed 7 November 2022.

39 Aditi Ramesh and Astha Kapoor, 'Principles for Revenue Models of Data Stewardship' (2020) <https://thedataeconomylab.com/2020/07/31/principles-for-revenue-models-of-data-stewardship/> accessed 4 November 2022.

40 Julian Trait, Open Data Manchester (n 13).

experimentative and haven't tested other revenue lines such as monetising data, charging commissions etc.[41] There is a need to solve many functional and moral questions to be an attractive alternative to the current ways in which the data economy is organised.

This flexibility, and context-specificity of models, along with legal and monetary open questions needs to be borne in mind as policy documents begin to recommend data stewardship.

Conclusion

As levels of digitisation and datafication of our lives and communities increase, the need to rethink how we govern data is imminent. Data stewardship offers an alternative to the current status quo - anchored in the ideas of social value of data, collective decision-making and participation, it aims to rethink our engagement both with data, and those that control it.

Data stewardship is a powerful idea, and the source of its influence is that it is community led, is bottom up and reflects the needs of the people it aims to serve, and not the technology companies, or governments who seek to collect and use it. But data stewardship is complex, as no two communities, or their needs are similar and therefore their imagination and pathways to justice are different. Further, jurisdictions and their ability and inclination to create space for radical instruments such as data stewards vary. The diversity of models, of governance and accountability frameworks, infrastructure, legal instruments are all critical to support and ensure that we are not pushing one type of stewardship over another, and create the space for communities to explore design choices that work for them.

To do this, more evidence from the ground is required - such that decisions of communities on questions of data are better understood and chronicled and serve as lessons, as this space grows. There are instruments such as sandboxes (as used by the Reserve Bank of India) that can be deployed to test the efficacy of stewardship in controlled circumstances. But beyond top-down interventions by the government, data stewardship needs to be anchored in communities that need help in understanding the value of their data, and the rights around that, and the need to reimagine the current structures of the data economy. It is through this bottom-up action that stewardship can be made a reality. Once the need is clearly established through community awareness and action, an investment in capacity is required. Thereafter, it is certain that regulatory changes that create space for innovations on data rights will be evolved.

---

41 Aditi Ramesh and Astha Kapoor (n 39).

Making Data Count - A Case for Developing Data Stewardship Models for the Indian Judiciary

Ameen Jauhar[1]

Introduction

The Indian judiciary has witnessed a steady augmentation of its technological infrastructure over the past two decades. Initiated under the E-Courts Mission Mode Project (e-Courts project), the drive to integrate conventional information and communication technologies with courts, has significantly transformed the justice system and justice delivery processes.[2] Recently, the e-committee of the Supreme Court of India has also put out a draft vision document stipulating details of the way forward for phase III of the e-Courts project.[3] One area that has become a talking point in this discourse of further deploying sophisticated and emerging technologies with the judiciary, is the role Big Data is likely to play in this process.[4]

The Indian justice system collects large amounts of personal and non-personal data (NPD) as part of its diurnal functions and routine processes. To give a broad overview, right from the stage of filing a case at the registry counter, to the actual litigation process, parties (and even lawyers) are required to furnish considerable sensitive information (sometimes even including religious affiliation), and other personal details like demographic information, and even visual or photographic ids.[5] In litigation, while filing affidavits, or to support their arguments, litigants are further required to furnish even more information either under a specific legislation, or to corroborate their factual arguments. For instance, in cases arising before family courts, sensitive information regarding the private family relationships, marital status, and even sexual orientation may become part of the record all of which is highly personal information. Resultantly, the Indian judiciary today is arguably collecting and archiving data as a significant institution. However, it is a warranted question - how is this data processed, shared, or utilised? Has the judiciary established a methodical and streamlined framework for data processing or not? Also, given that

---

1 Ameen is a Senior Resident Fellow at the Vidhi Centre for Legal Policy, leading its Centre for Applied Law & Tech Research (ALTR). He can be contacted at ameen.jauhar@vidhilegalpolicy.in.

2 E-Committee, Supreme Court of India, 'Digital Courts Vision & Roadmap: Phase III of the ECourts Project' <https://cdnbbsr.s3waas.gov.in/s388ef51f0bf911e452e8dbb1d807a81ab/uploads/2021/04/2021040344.pdf> accessed 31 October 2022.

3 ibid.

4 Ameen Jauhar, 'AI Innovation in Indian Judiciary a Distant Dream Without an Open Data Policy' (*Vidhi Centre for Legal Policy*, 14 April 2020) <https://vidhilegalpolicy.in/blog/ai-innovation-in-indian-judiciary-a-distant-dream-without-an-open-data-policy/> accessed 31 October 2022; Aakanksha Mishra and Siddharth Mandrekar Rao, 'Judicial Data Regulation' (DAKSH India 2021) <https://www.dakshindia.org/judicial-data-regulation/> accessed 31 October 2022.

5 For a general discussion on the data that the judiciary collects please see the E-Committee, Supreme Court of India, (n 2); Adrija Jayanthy and others, 'Open Courts in the Digital Age: A Prescription for an Open Data Policy' (VIdhi Centre for Legal Policy 2019) <https://vidhilegalpolicy.in/research/open-courts-in-the-digital-age/> accessed 31 October 2022.

the judiciary is a public institution, should this data collected and archived by it, be made publicly accessible?

These are some questions that emerge as soon as we begin recognising the copious datasets that the judiciary, through its operations, generates. As most scholarship on Big Data analytics suggests, it has the potential to spur significant tech innovation and provide disruptive and innovative solutions for existing institutions.[6] For the judiciary too, it is crucial to understand and devise nuanced systems which will allow it to streamline its data collection and processing in a responsible and safe manner. Such datasets can further facilitate social innovation aimed at improving access to justice, and the overall efficiency of courts in India. They can also create valuable "digital intelligence",[7] a practice of analyses and patterns emerging from datasets that can inform decision making, in this case, for better, data driven judicial reforms and policies.

To accomplish these twin objectives of accumulating digital intelligence, and sharing judicial data for social innovation, stewardship of such data is arguably an optimal mode of data governance and management. Data stewardship fundamentally creates a fiduciary role of an entity tasked with collection, collation and archiving, processing, and sharing of data in a responsible manner, furthering the public interest.[8] Also integral to data stewardship, is the idea of establishing a "data commons"[9] - a pool of openly accessible datasets, which are governed by an institutional structure, with the aim of furthering community and public interests.

For the judiciary, a data stewardship model would ensure two things. First, control over the data it accumulates and processes as an institution; and second, ensure that sharing and further usage of such data by third parties is dictated by a strong commitment to socially benevolent innovation rather than pure economic and commercial calculus.[10]

This essay will examine how data stewardship can be effectuated for the judiciary. This is a relatively novel idea even in general scholarship on data governance, let alone its application for a judicial structure. Hence, the essay aims to serve as a

---

6 See for a general discussion, James Manyika and others, 'Big Data: The next Frontier for Innovation, Competition, and Productivity' (McKinsey Global Institute 2011) <https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_exec_summary.pdf> accessed 31 October 2022.

7 Parminder Jeet Singh, 'Data and Digital Intelligence Commons (Making a Case for Their Community Ownership)' [2019] SSRN Electronic Journal <https://www.ssrn.com/abstract=3873169> accessed 31 October 2022.

8 Trishi Jindal and Aniruddh Nigam, 'Data Stewardship for Non-Personal Data in India' <https://vidhilegalpolicy.in/research/data-stewardship-for-non-personal-data-in-india/> accessed 31 October 2022.

9 Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (1st edn, Cambridge University Press 1990) <https://www.cambridge.org/core/product/identifier/9780511807763/type/book> accessed 31 October 2022; Brett M Frischmann, Michael J Madison and Katherine J Strandburg (eds), *Governing Knowledge Commons* (Oxford University Press 2014) <https://academic.oup.com/book/36261> accessed 31 October 2022. Knowledge commons have been created for the pooling and creation of genome research, free information.

10 Trishi Jindal and Aniruddh Nigam (n 8).

primer of some core concepts of data stewardship, and its application to the judiciary, particularly focusing on data trusts. It further aims to present some crucial factors that must be considered if this theoretical idea of judiciary's data trust is to be actually operationalised. For this, the present essay is broadly divided into three sections. The first part will delve deeper into the ongoing technological endeavours of the judiciary, and the role Big Data is likely to play in the future. The second part will elaborate on how to enable a robust and public centric data stewardship model for the judiciary and look at data trusts being an optimal entity to manifest this model of governance. The last section will list factors for consideration in order to establish a data trust for the Indian judiciary and follow the same with concluding remarks.

1. The Potential of Big Data for the Indian Judiciary

In the early 2000s, in alignment with the nationwide effort to promote e-governance, the e-committee of the Supreme Court of India was established.[11] The objective initially was to modernise Indian courts by automating certain processes. This was envisioned as a fruitful step to improve access to justice, make judicial administrative processes more streamlined, and overall augment the efficiency of courts. To this end, the first phase dedicated itself to providing the requisite computing infrastructure, especially at the district courts' level.[12] Some basic information from courts was also made available in electronic format, particularly case information (like case name and number, court room numbers, and online cause lists).[13] By 2014, when the second phase was to commence, focus had shifted from large scale procurement of hardware to creation of innovative and sophisticated software. This would include e-filing, e-payment, real time case updates,[14] creation of detailed websites for different courts, and automation of many other processes.[15]

Towards the end of the second phase in 2019, conversation also began exploring the potential of emerging technologies like AI.[16] This has been accompanied by an increasing advocacy for streamlining judicial data that gets accumulated across

---

11 Ministry of Law & Justice and (Department of Justice), 'Establishment of an E-Committee for Monitoring Use of Information Technology and Administrative Reforms in the Indian Judiciary' (8 December 2004) <https://main.sci.gov.in/pdf/ecommittee/ecommittee%20officeorder.pdf> accessed 31 October 2022.

12 E-Committee, Supreme Court of India, (n 2).

13 E-Committee Supreme Court of India, 'Policy and Action Plan Document Phase II of the E-Courts Project' <https://ecourts.gov.in/ecourts_home/static/manuals/PolicyActionPlanDocument-PhaseII-approved-08012014-indexed_Sign.pdf> accessed 31 October 2022.

14 Real time case updates refer to virtual display boards which replicate the physical display boards available in court premises. Through these updates it is easier to track the hearing of cases across different benches of the court, throughout the day, and determine when a certain matter is likely to come up for hearing.

15 E-Committee Supreme Court of India (n 13); E-Committee, Supreme Court of India, (n 2).

16 *Hon'ble Mr. Justice L.Nageswara Rao | Artificial Intelligence and the Law* (Shyam Padman Associates 2020) <https://www.youtube.com/watch?v=ZJsIQwPn5AU> accessed 31 October 2022; Neha Joshi, 'Artificial Intelligence Can Supplement but Not Supplant a Judge: CJI SA Bobde' (*Bar and Bench*, 16 April 2021) <https://www.barandbench.com/news/litigation/artificial-intelligence-supplement-replace-judge-chief-justice-sa-bobde> accessed 31 October 2022; Ameen Jauhar, 'Can Artificial Intelligence Help Reform Indian Courts? | Opinion' *Hindustan Times* (28 November 2019) <https://www.hindustantimes.com/analysis/can-artificial-intelligence-help-reform-indian-courts-opinion/story-JuJBA5lLJyLZNCPWSGs9fJ.html> accessed 31 October 2022.

all tiers of the judiciary.[17] Specifically, over the last two years, through an internal AI committee constituted by the Supreme Court of India, two algorithmic tools have been piloted. The first is a neural translation tool called SUVAS that allows the translation of judgments and orders from English to niche Indic vernaculars and vice-versa.[18] The second is SUPACE, a case management algorithm that also has features of case query and analytics.[19] Additionally, there has been a growing conversation around legal technology and justice stack, which focus on imbibing tech solutions in a platform and holistic manner, rather than through piecemeal pilots.[20] Both these pilots involve AI techniques like machine learning and natural language processing. Machine learning (ML) has become a common technique among AI specialists to design task specific algorithms which cannot mimic all aspects of human cognition (or general AI) but are able to perform their dedicated tasks at a much faster and efficient rate. Similarly, natural language processing (NLP), is another form of creating intelligent algorithms reliant on large and expanding data corpses to understand human languages and interact with them accordingly.[21] All these discussions focusing on intelligent algorithms, have a common denominator - the necessity for unbridled access to large judicial datasets which will be pivotal in developing and training the underlying algorithms.[22]

While open access to judicial data is a *sine qua non* for continuous innovation of legal tech, it also has the potential to aid in better informed judicial policies and reforms.[23] A crucial development of the digital economy has been the increasing amount of "digital intelligence" that emerges from data analytics.[24] Simply put, digital intelligence refers to sophisticated analytics and patternisation in datasets that can (ironically) be identified by algorithms trained to conduct such analyses. This, in a commercial environment, has proven to be pivotal for companies to amend their internal practices, corporate operations, consumer targeting and marketing strategies, to name a few areas of impact.[25]

---

17 Adrija Jayanthy and others (n 5); Ameen Jauhar (n 4).

18 Supreme Court of India, 'Press Release | 25 November 2019' <https://main.sci.gov.in/pdf/Press/press%20release%20for%20law%20day%20celebratoin.pdf> accessed 31 October 2022.

19 Shanthi S, 'Behind SUPACE: The AI Portal Of The Supreme Court of India' [2021] *Analytics India Magazine* <https://analyticsindiamag.com/behind-supace-the-ai-portal-of-the-supreme-court-of-india/> accessed 31 October 2022.

20 Amitabh Kant, Preeti Syal, and Desh Gaurav Sekhri, 'Time for a Justice Stack' *Financial Express* (14 July 2021) <https://www.financialexpress.com/opinion/time-for-a-justice-stack/2289629/> accessed 31 October 2022; The NITI Aayog Expert Committee on ODR, 'Designing the Future of Dispute Resolution: The ODR Policy Plan for India' (2020) Draft for Discussion <https://www.thehinducentre.com/publications/policy-watch/article34777275.ece/binary/Draft-ODR-Report-NITI-Aayog-Committee.pdf> accessed 31 October 2022.

21 Natural Language Processing, usually shortened as NLP, is a branch of artificial intelligence that deals with the interaction between computers and humans using the natural language. The ultimate objective of NLP is to read, decipher, understand, and make sense of the human languages in a manner that is valuable. Most NLP techniques rely on machine learning to derive meaning from human languages. Michael J Garbade, 'A Simple Introduction to Natural Language Processing' (*Medium*, 15 October 2018) <https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32> accessed 31 October 2022.

22 Ameen Jauhar (n 4).

23 Adrija Jayanthy and others (n 5).

24 Singh (n 7).

25 ibid.

An unconventional paper in legal scholarship, discussing the benefits of a data driven approach to judicial reforms was published some years back by constitutional law expert and present vice-chancellor of India's premiere law school, Prof. Sudhir Krishnaswamy.[26] He and the other authors discussed how better empirical evidence on the judiciary was needed to understand some perennial issues (like backlog and pendency) with the Indian judiciary in greater depth, than the current surface level methodology. The present author has also worked on this issue of a more empirical evidence-based approach to designing judicial reforms in his own dissertation thesis. This idea, however, is not merely entrenched in academic brainstorming. The Department of Justice of the Indian government's Ministry of Law & Justice, established the National Judicial Data Grid, under the larger e-Courts project, to collect more detailed statistical data on cases and pendency across district and high courts of India.[27] The Supreme Court of India also periodically publishes pendency statistics across high courts and district courts.[28] However, both these manifest, still a shallow appreciation of the potential of data analytics and the role it can play in better identification of problems of the judiciary and creating targeted solutions for them.

The aforementioned idea of digital intelligence can truly transform this existing myopia in judicial reforms. Assuming the creation of openly accessible, machine-readable judicial datasets, presents fantastic opportunities for designing special analytical algorithms which can take over this role of data syntheses from human operators. Such a tool can prove to be futuristic and supplement judges in the constant endeavours of trying to improve judicial access. For instance, the constitution of specialised benches is a debatable reform that is adopted by the judiciary to address specific backlogs. Typically, it is contingent on factors like excessive backlog of a specific type of litigation, or the urgency to dispose of specific cases, based on extraneous considerations. Algorithms typically tasked with operational management functions can be a tech intervention that aids in a more viable constitution of such specialised benches, arguably backed in a more evidence-based approach rather than anecdotal.

Another good example of such application could be case query and synthesis tools which can furnish pointed responses to queries posed by judges about the factual matrix of such dispute, either during the litigation proceeding, or even prior to it once the pleadings have been filed. This could ease the time spent in perusal of voluminous documents while adjudicating disputes and can be supplemented with oral arguments. While pendency and backlog have been the forefront challenges, judicial

---

26 Sudhir Krishnaswamy, Sindhu K Sivakumar and Shishir Bail, 'Legal and Judicial Reform in India: A Call for Systemic and Empirical Approaches' (2014) 2 Journal of National Law University Delhi 1 <http://journals.sagepub.com/doi/10.1177/2277401720140101> accessed 31 October 2022.

27 'National Judicial Data Grid' <https://njdg.ecourts.gov.in/njdgnew/index.php> accessed 31 October 2022; Adrija Jayanthy and others (n 5).

28 'Supreme Court of India | Publication | Annual Reports' <https://main.sci.gov.in/publication> accessed 31 October 2022.

reforms today discuss an array of problems ranging from poor infrastructure,[29] to the lacking diversity in the judicial class.[30] Not all issues are a technological problem; yet a more nuanced understanding of each issue through intelligent analytics, is arguably a better approach to policy making than the current practice driven by intuition and anecdotal evidence. Management and administrative algorithms can be designed to provide more detailed insights into why issues like backlog of cases, continue to persist despite a steady increase in judicial capacity and budgeting.

For both these avenues of innovative tech interventions, as well the development of a consistent and significant body of digital intelligence, judicial data needs to be streamlined. It brings us to the issue of how such data can be processed safely, preserving individual privacy (if personal information is also being processed), and community interest (in the case of NPD). Furthermore, given the inherent objective is to improve the judiciary as an institution, how can the public interest be safeguarded even when such judicial data is made accessible to third parties. The following section will be delving into these systemic considerations for establishing a data stewardship model for the Indian judiciary.

2. Stewardship and Management of Judicial Data in India

With the increasing process of digitisation, the judiciary in India is collecting significant data. Broadly speaking, this could be classified into three categories - first, personal data that may be furnished to courts either at the time of filing of a case, or during the proceedings voluntarily or by mandate of law (for eg: furnishing personal information such as name, age, gender, and address when a party stipulates any facts on an affidavit). The courts also collect information that can be categorised as NPD, or which does not result in the identification of an individual.[31] A good example of NPD would be the collation of case statics which may not result in personal identification but do serve a significant purpose in terms of developing data-driven strategies for judicial reform. A key example here would be the constitution of specialised benches which typically emerges from determining pendency of specific types of cases.

For the purpose of this essay, the author will be examining the second category of information collected, i.e., NPD. For personal data, given the data principal being the seminal source of the same, there is little debate on the interest she has in

---

29 Special Correspondent, 'Judicial Infrastructure Key for Improving Access to Justice, Says CJI' *The Hindu* (23 October 2021) <https://www.thehindu.com/news/national/other-states/cji-ramana-rues-ad-hoc-unplanned-improvement-and-maintenance-of-judicial-infrastructure/article37136774.ece> accessed 31 October 2022.

30 Sumathi Chandrashekaran and others, 'Breaking through the Old Boys' Club : The Rise of Women in the Lower Judiciary' 55 Economic and Political Weekly <https://www.epw.in/journal/2020/4/special-articles/breaking-through-old-boys%E2%80%99-club.html> accessed 31 October 2022.

31 The usual understanding of NPD is a collective reference to any mass-data which is not personal in nature. This includes anonymised datasets but is not limited to them. For example, see Joint Parliamentary Committee, *Report of the JPC on the Personal Data Protection Bill, 2019* (Recommendation No. 26, 2021).

exercising autonomy over such information.[32] However, with NPD this control of the individual(s) or the community from where the same may emanate, becomes debatable. What has further exacerbated this idea of "community interest" in NPD is how the data economy seems to be largely driven by for-profit corporations who have harvested large amounts of such datasets to generate crucial digital intelligence pivotal in their commercial decision making.[33] A direct consequence of this is great reticence within the judiciary regarding the sharing of NPD, or any judicial data that it collects, with third parties. The impression is that any such access will inevitably yield expensive, proprietary protected technologies, in lieu of open-source ones. This will arguably impede public interest driven and promote profiteering as the predominant interest.[34]

Stemming from this suspicion of commercial undertakings being involved in the development of technologies for judiciary, is the idea that oversight is not necessary but unavoidable. The manner in which the e-Courts project has so far developed is a manifestation of what the judiciary deems as adequate oversight. Even with respect to data sharing, it is expected that the judiciary will want any such frameworks to be developed under its auspices with it retaining final say over the same. The author proposes that a data stewardship model for the judiciary must comply with these two aspects, namely promoting social interest and ensuring adequate control of the judiciary over any technology designed and deployed.

A common form of data stewardship, especially common for open access to datasets intended for tech innovation, is that of a data exchange. This typically involves a *laissez faire* approach where datasets are uploaded to exchange and are accessible to members or users of that platform.[35] For instance, the Telangana government is presently setting up a data exchange for agricultural datasets which will contain information around weather forecasts, cropping patterns, soil quality, etc.[36] While such an exchange is a tried and tested facilitator of interactions between the data sharers and data users, it is important to highlight that there is little to none oversight involved. The structure is geared towards easy access of data for third party users, which may or may not be anchored in public welfare. Also, such an entity lacks the institutional governance that would be crucial in establishing the judiciary's supremacy in such an arrangement of data sharing.

---

32 *K.S. Puttaswamy v. Union of India* (2017) 10 SCC 1, especially [142] (plurality opinion of Justice Chandrachud); Vrinda Bhandari and others, 'An Analysis of Puttaswamy: The Supreme Court's Privacy Verdict' [2017] IndraStra Global 6 <https://www.ssoar.info/ssoar/handle/document/54766> accessed 31 October 2022.

33 Bertin Martens, 'How Online Platforms Challenge Traditional Views of the Firm' (August 2016) <http://blogs.oii.ox.ac.uk/ipp-conference/2016/programme-2016/track-c-markets-and-labour/government-regulation-of-platforms/bertin-martens-how-online-platforms.html> accessed 31 October 2022.

34 *Justice Madan Lokur | Roadmap for Establishing Virtual Courts in India* (Vidhi Centre for Legal Policy 2020) <https://www.youtube.com/watch?v=ITSfonSVMZk> accessed 31 October 2022.

35 Trishi Jindal and Aniruddh Nigam (n 8).

36 'Telangana Open Data Portal' <https://data.telangana.gov.in/search/field_topic/agriculture-36> accessed 3 November 2022.

Another model of data stewardship is one of a data cooperative, which is a shift from a free market, unregulated approach, to a more collective pooling of data resources. This collective allows some amount of control on how the data is collated, processed, and shared, but for any entity or individual not part of the collective, it is impossible to influence these decisions. While such a cooperative can further public interest, if the same is amenable to all members, it vitiates the paramount position of the judiciary that is likely to be non-negotiable.

Data trusts are a third format of data stewardship that have increasingly found their way into the discourse of data governance at large.[37] Drawing from the core ideas of legal trusts, this structure formulates a fiduciary governance model where a group of trustees are posited as custodians of public interest.[38] In such a scenario, the trustee(s) are required to be independent of the data principals, data collectors, data sharers, and potential end users.[39] In order to discharge its fiduciary functions, the trustees are at liberty to adopt protocols and procedures, including setting out terms of license for the access granted to third parties. The trust model is one which effectively bridges the open access ideology of data exchanges, with an institutional framework to establish an effective yet flexible governance framework.

In the context of the Indian judiciary, a data trust is most likely to accomplish the twin objectives of data control and socially benevolent innovation stated previously. The question remains on how such a trust is to be established as a legal entity. Legal trusts in Indian law are mostly private in nature, where the trustees act as custodians of a certain tangible asset for certain private individuals or entities, rather than the public at large. More importantly, a data trust, while deriving fiduciary principles, is arguably a novel legal entity. This is primarily for two reasons - first, data, including NPD, is highly contested as a tangible asset or property;[40] and secondly, there is a lack of any jurisprudential or legal basis of data being a property that is transferable to a beneficiary, under Indian law.[41] Therefore, data trust needs the institutional trappings of an entity which can allow the performance of fiduciary duties, while entrusting the said trust with the comprehensive rights over NPD emanating from the judiciary.

---

37 Anouk Ruhaak, 'Data Trusts: Why, What and How?' (*Medium*, 13 November 2019) <https://medium.com/@anoukruhaak/data-trusts-why-what-and-how-a8b53b53d34> accessed 31 October 2022; Bianca Wylie and Sean Martin McDonald, 'What Is a Data Trust? - Centre for International Governance Innovation' (*Centre for International Governance Innovation*, 9 October 2018) <https://www.cigionline.org/articles/what-data-trust/> accessed 31 October 2022.u

38 Ruhaak (n 37).

39 Open Data Institute, 'ODI Report: "Data Trusts: Lessons from Three Pilots" #EXTERNAL' (*Google Docs*) <https://docs.google.com/document/d/118RqyUAWP3WIyyCO4iLUT30ObnYJGibEhspr2v87jg/edit?usp=embed_facebook> accessed 31 October 2022.

40 Lalit Panda, 'The Hybridisation of Property, Liability and Inalienability in Data Protection' (2020) 4 Journal of Intellectual Property Studies 18 <https://journalofipstudies.files.wordpress.com/2022/10/hybrid-prop-1.pdf> accessed 31 October 2022.

41 Trishi Jindal and Aniruddh Nigam (n 8).

In this background, the next section will conclude by giving some ideas on how such a data trust for the judiciary can be conceptualised within the trappings of a not-for-profit company under Section 8 of the Companies Act, 2013.

Way Forward - Conceptualising a Data Trust for the Indian Judiciary

According to the Open Data Institute in the United Kingdom, the governing institution is crucial to the overall data infrastructure for data sharing.[42] For the proposed stewardship of judicial NPD, it is crucial that this institutional framework is established thoughtfully. Keeping this in mind, the following recommendations are proposed regarding a data trust for the Indian judiciary:

A. ***Balancing judicial oversight and independence of the trust*** - By definition and function, a data steward is required to be independent of data collectors, sharers or users.[43] However, for the judiciary, given the novelty of this form of data governance, it would be advantageous to establish some confidence building measures. One such measure is to have adequate representation of the judiciary in the trusteeship model. In this regard, given the normal inclination of courts and judges to prescribe oversight mechanisms, it will be critical to create protocols where the judiciary can voice its points without undermining the independence of the stewardship model.

B. ***Establishment of a dedicated entity*** - To further secure the buy-in of the judiciary to the idea of a data trust, it will be useful incorporating the same as a not-for-profit entity. A section 8 company within the Companies Act, 2013, can serve as a useful vehicle which creates a permanent and dedicated institution for handling everyday affairs of judicial data management. Within the company, its incorporation documents can clearly stipulate its objectives and mission of furthering access to NPD aggregated by the judiciary, driven by public interest and socially beneficial tech innovation.

C. ***Creating institutional and technological layers*** - Fundamentally, the data trust will require a technological layer, and an institutional layer. The former could be established within the company as different committees which oversee functional aspects of diurnal governance. The latter could be set up as a digital platform wherein datasets can be collated, archived, processed, and structured for sharing.

---

42 Open Data Institute (n 39).

43 Trishi Jindal and Aniruddh Nigam (n 8).

D. *Determining institutional frameworks to discharge fiduciary responsibilities* - A crucial reason for opting a data trust model of stewardship is the fact that in theory, it is most equipped to discharge the fiduciary obligations. For the judiciary, the preservation of public interest is inherent to its constitutional role and positioning, and thus, the discharge of fiduciary responsibilities even with the collection and sharing of judicial data will be a priority. Any data trust for the judiciary must have adequate internal frameworks to ensure this role is not merely titular but carried out meaningfully. The charter documents of such an organization will need to establish clear roles and legally enforceable obligations on the part of the stewards, to undertake these fiduciary functions.

E. *Participatory decision-making and engagement with different stakeholders* - The data trust model is not envisioned as a top-down governance framework where once data is stored, it goes beyond the control of data principals or other stakeholders. The stewards act as mediators of competing interests in a data economy, and as such must ensure continuous engagement with these different stakeholders. A proposed data trust for the judiciary must also be conceptualised in this manner - it should devise engagement protocols and mechanisms wherein the different stakeholders feel utility and trust the steward's neutrality and fiduciary nature.

The idea of a data trust proposed in this essay is novel and certainly requires more unpacking. There is also an argument that a data trust may require a bespoke legislation to effectively perform the fiduciary duties and establish itself as a robust stewardship model.[44] That said, as this essay demonstrates, there is a pressing need for better data governance and management frameworks for the judiciary. It's tall order of deploying ever more sophisticated technologies and implementing impactful reforms require a better recognition of the role data collection and sharing is going to play in these processes. The hope of the author is that this introductory document can serve as a seminal point of building this discourse for the Indian judiciary.

44 Open Data Institute (n 39).

Emotion Recognition and the Limits of Data Protection

Vidushi Marda[1]

Introduction

In 2021, Uttar Pradesh police announced plans of using facial recognition cameras to detect 'distressed' women on the streets of Lucknow as part of their Safe City Project.[2] The multi-national company Mettl (based in India and the US) offers customers access to a 'dark personality inventory' which claims to measure negative personality traits (namely opportunism, self-obsession, insensitivity, temperamental, impulsiveness and thrill-seeking) in potential hires and existing employees.[3] Bangalore-based company Entropik Technologies offers a suite of commercial products that claim to infer emotions from facial expression, eye gaze, vocal tonality and brainwaves.[4]

Emotion recognition systems like these three-use machine learning to purportedly infer a person's inner emotional state and classify them into categories like fear, anger, surprise, happiness, etc. This represents an evolution in biometric technologies, from identifying who a person is (as is the case with facial recognition), to determining what a person apparently feels. Emotion recognition technologies make inferences from various forms of input data, including facial expressions, vocal tone, gait, physiological signals, among others.[5]

These technologies fall under what scholars Luke Stark and Jevan Hutson have recently termed "Physiognomic AI", which they define as "*The practice of using computer software and related systems to infer or create hierarchies of an individual's body composition, protected class status, perceived character, capabilities, and future social outcomes based on their physical or behavioral characteristics*".[6] Physiognomic AI applications have witnessed a resurgence in recent years, and are now claimed to be able to infer a person's characteristics, (like political leanings and sexual orientation),[7] future

---

1 Vidushi is a Senior Programme Officer at ARTICLE 19. She can be reached via her website https://vidushimarda.com.

2 Pathikrit Chakraborty, 'UP Cops to Use AI to Read Faces and Help Women in Distress' *Times of India* (Jnuary 2021) <https://timesofindia.indiatimes.com/india/up-cops-to-use-ai-to-read-faces-and-help-women-in-distress/articleshow/80396572.cms>.

3 'Mettl Dark Personality Inventory' <https://mettl.com/dark-personality-inventory/>.

4 'Entropik Tech' <https://entropiktech.com/>.

5 Vidushi Marda and Shazeda Ahmed, 'Emotional Entanglement: China's Emotion Recognition Market and Its Implications for Human Rights' (Article 19 2021) <https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf>.

6 Luke Stark and Jevan Hutson, 'Physiognomic Artificial Intelligence' [2021] Fordham Intellectual Property, Media & Entertainment Law Journal, Forthcoming <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3927300>.

7 Yilun Wang and Michal Kosinski, 'Deep Neural Networks Are More Accurate than Humans at Detecting Sexual Orientation from Facial Images' [2017] PsyArXiv.

behavior (eg: predicting criminality, trustworthiness as an employee),[8] and inner emotional state (eg: detecting deception, fear, anger across a number of use cases).[9]

Face-based emotion recognition is a particularly popular application of physiognomic AI given the readily available pre-existing infrastructure facilitating face recognition systems. Emotion recognition is not even necessarily considered explicitly in existing policy proposals, but rather assumes the role of the next logical step in the trajectory of biometric technologies, as seen in the Lucknow Safe City tender.[10]

Tangible use cases exist across the world. *iBorderCtrl* was a trial program in Europe which used emotion recognition to detect deception at immigration checkpoints,[11] ViaQuatro in Brazil fitted in cameras on the Sao Paulo subway to detect emotion, gender, and age of passersby to better serve ads.[12] China is home to a burgeoning market for emotion recognition applications that are currently being trialed and used for a number of use cases including public security, education and driving safety. One emotion recognition company in China even claimed that emotion recognition heralds the phase of "Biometrics 3.0", with fingerprints and facial recognition being the preceding two stages.[13]

But experts disagree on whether emotion recognition can work in the first place. A significant body of scientific work argues that emotion recognition is based on junk science, even as companies and governments turn towards deploying it. Discredited for centuries, it draws from a branch of pseudoscience called physiognomy that studies a person's facial features or shape of the body in relation to their character.[14] Physiognomic thought was most prominently associated with the Nazi racial purity agenda in recent decades, but has nevertheless endured the test of time. Even so, as evidenced by the examples above, academic research reproduces and builds on these ideas; companies market emotion recognition systems and also use them to surveil employees and track consumers; and state actors procure this technology to enhance public safety, security and law & order.

---

8 Xiaolin Wu and Xi Zhang, 'Automated Inference on Criminality Using Face Images' [2016] Arxiv <https://arxiv.org/abs/1611.04135v1?utm_campaign=Tech%2520Policy%2520Daily&utm_medium=email&utm_source=Revue%2520newsletter>.

9 Vidushi Marda and Shazeda Ahmed (n 5).

10 The Lucknow Safe City project published a "Request for Proposal for Selection of System Integrator for Design, Implementation and Maintenance of Integrated Smart Control Room (ISCR)" in June 2021. At the time of writing this piece, no tender award has been announced. While Staqu Technologies works with the Uttar Pradesh government on general surveillance (see Internet Freedom Foundation's Panoptic Tracker: https://panoptic.in/uttar-pradesh/FRT-000004) there is no indication in the public domain of which entities this particular safe city tender has been awarded to.

11 Javier Sanchez Monedero and Lina Dencik, 'The Politics of Deceptive Borders: Biomarkers of Deceit and the Case of IBorderCtrl' [2019] Computers and Society <https://arxiv.org/pdf/1911.09156.pdf>.

12 Veronica Arroyo and Daniel Leufer, 'Facial Recognition on Trial: Emotion and Gender "Detection" under Scrutiny in a Court Case in Brazil' (*AccessNow*, 29 June 2020) <https://www.accessnow.org/facial-recognition-on-trial-emotion-and-gender-detection-under-scrutiny-in-a-court-case-in-brazil/>.

13 Vidushi Marda and Shazeda Ahmed (n 5).

14 Luke Stark and Jevan Hutson (n 6).

This essay will scrutinise face-based emotion recognition technologies in the context of data governance in India. I argue that any attempts to oversee data collection, processing and sharing in the context of emotion recognition systems are a feeble and ineffective way of regulating this technology. I demonstrate why the only feasible data governance approach with respect to emotion recognition technology is to reject the design, development, testing and deployment of these systems altogether.

It is crucial to do so at this juncture, for a number of reasons. First, the political appetite for biometric technologies has grown significantly across the world, including in India, and it is only a matter of time before the attention and interest State actors allocate to emotion recognition reaches a crescendo, particularly as wider policy agendas like 'smart cities' gain momentum. Secondly, working on these issues now, i.e. before these technologies are ubiquitous and invested in, provides academic and civil society actors with the time necessary to scrutinise and critique the existence of these technologies while offering constructive solutions for the way forward. And finally, a number of countries including India are in the midst of regulatory developments in the context of AI and data protection, and the unique challenges posed by physiognomic AI technologies like emotion recognition must be rigorously studied and engaged with.

I make this argument as follows. The next section will discuss the legacy and nature of emotion recognition technology. Section III will reflect on the current state and limitations of a data governance framework in India, and demonstrate why the only way it can effectively regulate harms is by refusing to engage with these technologies at all. Section IV will conclude.

1. Emotion Recognition: A Short Primer on Underlying Assumptions and History

Face-based emotion recognition systems use machine learning to (i) detect a face, (ii) detect emotional expression, and finally, (iii) classify such expression against an emotion.[15]

Commercial emotion recognition technologies are largely based on psychologist Paul Ekman's Basic Emotion Theory (BET) which argues that "*there should be bodily signatures for each basic emotion consisting of highly correlated and emotion-specific changes at the level of facial expressions, autonomic changes and preset and learned actions*."[16]

---

15 To understand steps involved in face recognition particularly pertaining to (i), see Stan Z. Li and Anil K. Jain, 'Overview of Facial Recognition Process' [2011] Handbook of Face Recognition; Dilbag Singh, 'Human Emotion Recognition Systems' (2012) 4 International Journal of Image, Graphics and Signal Processing; Jacintha V. and others, 'A Review on Facial Emotion Recognition Techniques' International Conference on Communication and Signal Processing, IEEE Explore <https://ieeexplore.ieee.org/document/8698067>.

16 Andrea Scarantino and Ronald de Sousa, 'Emotion' [2021] The Stanford Encyclopedia of Philosophy; Also see Paul Ekman and Wallace V. Friesen, 'Nonverbal Leakage and Clues to Deception' 32 Psychiatry 88.

In other words, inherent in these systems are a few assumptions: that these facial expressions are universal and can be classified into discrete categories; that they are true and involuntarily "leak" onto faces; and finally, that there is a reliable link between an individual's inner emotional state and their facial expressions. While these are the building blocks at the centre of a rapidly growing global industry, none of these assumptions stand the test of scientific scrutiny.

1.1. On Universality

Paul Ekman led a group of scientists in the 1960s in an effort to demonstrate that a few 'basic emotions' could be inferred from facial expressions across the world, i.e. that some facial expressions were universal.[17] As part of this research, subjects from various parts of the world were shown photographs of facial expressions, and asked to classify them against a set of words or stories that best described the picture according to them. Experiments were conducted with individuals from five "literate" cultures (namely, Argentina, Brazil, China, Japan and the United States) and two "preliterate" cultures from New Guinea (namely, the Fore linguistic cultural group and the Grand Valley Dani). Ekman's findings indicate that subjects were able to successfully classify facial expressions, which led his work to conclude that some facial expressions were indeed universal, i.e. for some facial expressions, muscular movement is associated with certain emotions through inheritance.[18]

But Ekman's work and his findings have been refuted since the time of their publication. In 1975, cultural anthropologist Margaret Mead reviewed Ekman's findings and methodology and found it to be "*an example of the appalling state of the human sciences*" given its failure to consider the disciplines of anthropology, psychiatry, sociology among others.[19] Ekman's 'natural kind' view of emotions - as something that is biologically inherited independent of our experiences and culture - has been refuted in recent years, prominently by psychologist Lisa Feldman Barrett who states, *"the natural-kind view has outlived its scientific value, and now presents a major obstacle to understanding what emotions are and how they work"*.[20] Barrett, in turn, proposes that emotions are constructed, and learned through human experience and social, cultural realities.

---

17 Paul Ekman and Wallace V. Friesen, 'Constants across Cultures in the Face and Emotion' (1971) 17 Journal of Personality and Social Psychology 124; also see Paul Ekman, Richard Sorenson, and Wallace V. Frisen, 'Pan-Cultural Elements in Facial Displays of Emotion' (1969) 164 Science <https://www.science.org/doi/10.1126/science.164.3875.86>.

18 Paul Ekman, 'Universals and Cultural Differences in Facial Expressions of Emotion' (1971) 19 Nebraska Symposium on Motivation <https://1ammce38pkj41n8xkp1iocwe-wpengine.netdna-ssl.com/wp-content/uploads/2013/07/Universals-And-Cultural-Differences-In-Facial-Expressions-Of.pdf>.

19 Margaret Mead, '"Margaret Mead Calls "Discipline-Centric" Approach to Research an "Example of the Appalling State of the Human Sciences"' [1975] Journal of Communication <https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1460-2466.1975.tb00574.x>.

20 Lisa Feldman Barrett, 'Are Emotions Natural Kinds' (2006) 1 Perspectives on Psychological Science 28.

Experts have studied Ekman's methodology of using a certain number of predetermined words to describe emotions and found that it primed subjects towards the correct answer, in turn skewing results.[21] When psychologist Lisa Feldman Barret and her team conducted similar experiments to Ekman but did not provide these preselected categories, they found that subjects' performance plummeted significantly, finding that *"their performance was comparable to that of people suffering from semantic dementia, who can distinguish positive from negative emotions in faces, but nothing finer."*[22]

Multiple additional studies have also refuted the idea of universal expressions. In 1995, psychologist J.A Russell demonstrated that emotions like anger and sadness are not pancultural and neither are they precisely conveyed across cultures. Russell suggested that at the very least there is a 'minimum universality' - i.e. "*people everywhere can infer something about others from their facial behavior*".[23] This resonates with an instinctive understanding of facial expressions and how people understand the world, while at the same time demonstrating that the assumption of universal expressions is a faulty one. In 2016, a group of researchers led by psychologist Carlos Crivelli found that a gasping face was interpreted as conveying fear and submission by Western adolescents, and as conveying anger and threat by adolescents from a Melanesian society isolated - culturally and visually - from the West.[24] In 2018, researchers from the University of Glasgow found that facial expressions of pain and orgasms are represented distinctly across cultures.[25]

1.2. On the Link Between External Facial Expressions and Internal Emotional States

BET assumes that inner emotional states can be inferred from external markers such as facial expression. This again, is scientifically suspect. Ekman along with Wallace V. Friesen proposed that microexpressions - facial expressions that occur within a fraction of a second - are involuntary and 'leak' onto faces to reveal a person's true emotions, before the individual is able to control their expression in response to stimuli that induce emotion.[26] To codify microexpressions and what they mean, Ekman and Friesen developed the Facial

---

21 Lisa Feldman Barrett, 'What Faces Can't Tell Us' *The New York Times* (28 February 2014) <https://www.nytimes.com/2014/03/02/opinion/sunday/what-faces-cant-tell-us.html>.

22 ibid; Kristen A Lindquist and others, 'Language and the Perception of Emotion' [2006] Emotion <https://pubmed.ncbi.nlm.nih.gov/16637756/>; Maria Gendron and others, 'Emotion Words Shape Emotion Percepts' (2012) 12 Emotion 314.

23 James A. Russell, 'Facial Expressions of Emotion: What Lies Beyond Minimum Universality?' (1995) 118 Psychological Bulletin 379.

24 Carlos Crivelli and others, 'The Fear Gasping Face as a Threat Display in Melanesian Society' (2016) 113 Proceedings of the National Academy of Sciences 12403.

25 Chaona Chen and others, 'Distinct Facial Expressions Represent Pain and Pleasure across Cultures' (2018) 115 Proceedings of the National Academy of Sciences <https://www.pnas.org/content/115/43/E10013>.

26 Paul Ekman and Wallace V. Friesen (n 16); also see Paul Ekman Group, 'What Are Micro Expressions?' <https://www.paulekman.com/resources/micro-expressions/>.

Action Coding System (FACS) in 1978 to enable the analysis and classification of facial muscle movements and in turn, emotions.[27] FACS continues to be a foundational element of emotion expression techniques to this day.

Like universality of emotions, microexpressions have been discredited for a number of reasons. In 2019, a panel of experts reviewed over a 1,000 scientific papers that explored the link between facial expressions and emotional states. They found, "*very little is known about how and why certain facial movements express instances of emotion, particularly at a level of detail sufficient for such conclusions to be used in important, real-world applications. Efforts to simply 'read out' people's internal states from an analysis of their facial movements alone, without considering various aspects of context, are at best incomplete and at worst entirely lack validity, no matter how sophisticated the computational algorithms*".[28]

Similarly, scholars have demonstrated that facial expressions are not solely related to emotional states, and have multiple causes and meanings.[29] People can feel multiple emotions at the same time and exhibit one expression, or a single emotion can inspire multiple expressions, depending on the individual's knowledge and experience about their affective state.[30] As Russell states in his argument for minimum universality of emotions, "*emotions can occur without facial expressions, and facial expressions can occur without emotions*".[31] To treat them as proxies for one another is to wholly reject overwhelming scientific evidence.

The practice of inferring internal states from external markers draws from physiognomic thought. While the practice of physiognomy can be traced back to ancient Greece and India, it had lost its popularity by the end of the 17th century.[32] Physiognomy in its present day form can be most directly associated with the writings of Johann Caspar Lavater, an 18th century pastor from Zurich who wrote a four part treatise on physiognomy which he claimed provided "universal axioms and incontenstible principles" which experts like Alexander Todorov refute, stating that Lavater's evidence *"came from counterfactual statements peppered with what now would be considered blatantly racist beliefs"*.[33]

---

27 Paul Ekman and Wallace V. Friesen, 'Measuring Facial Movement' (1976) 1 Environmental Psychology and Non Verbal Behavior <https://www.paulekman.com/wp-content/uploads/2013/07/Measuring-Facial-Movement.pdf>.

28 Lisa Feldman Barrett and others, 'Emotional Expressions Reconsidered: Challenges to Inferring Emotion from Human Facial Movements' 20 Psychological Science in the Public Interest <https://journals.sagepub.com/doi/10.1177/1529100619832930>; J.A. Russell, 'Is There Universal Recognition of Emotion from Facial Expression? A Review of the Cross-Cultural Studies' (1994) 115 102.

29 José-Miguel Fernandez-Dols and Carlos Crivelli, 'Emotion and Expression: Naturalistic Studies' (2013) 5 Emotion Review 24.

30 Lisa Feldman Barrett, 'Solving the Emotion Paradox: Categorization and the Experience of Emotion' (2006) 10 Personality and Social Psychology Review 20.

31 James A. Russell (n 23).

32 Kenneth Zysk, 'Greek and Indian Physiognomics' (2018) 138 Journal of the American Oriental Society 313.

33 Alexander Todorov, *Face Value: The Irresistible Influence of First Impressions"* (Princeton University Press 2017).

Even as Lavater's work was discredited shortly after its popularity peaked, the practice of physiognomy diffused through Europe in the 19th century.[34] The founding of criminal anthropology and eugenics (by Cesare Lombroso and Francis Galton respectively) witnessed physiognomic ideas being adopted as ground truth. Lombroso argued that *"thieves are notable for their ... small wandering eyes that are oblique in form, thick and close eyebrows, distorted or squashed noses, thin beards and hair, and sloping foreheads"*. He even argued that children who exhibited "*the smallness of the head, and the exaggerated size of the face*" would have "*scholastic and disciplinary shortcomings*" and should thus be separated from their "*better endowed companions*".[35] Galton on the other hand, believed that each race had a central type, and those deviating from this "ideal form" should be restricted from breeding. As some experts state, "*When put into practice, the pseudoscience of physiognomy becomes the pseudoscience of scientific racism*".[36]

The state of scientific consensus should make clear the futility of investing time, energy and resources in procuring emotion recognition technologies. And yet, it is increasingly being used in critical decision making - as an investigatory tool by law enforcement agencies to detect deception in individuals being interrogated, for assessing candidates at job interviews, for monitoring prisons, for surveilling borders, and as a tool for workplace surveillance across sectors. In reality, this has not stopped the steady growth of the emotion recognition market because of a number of factors. Firstly, academics, companies and State actors that are enthusiastic about this technology around the world embed and perpetuate lofty claims about what the technology can do, actively ignoring evidence to the contrary.[37] Assumptions transfer from academic papers to policy documents to marketing materials, lending weight to assertions with every turn, and entities to hold these claims to account, like civil society, or regulators are either not invited into deliberations, or are gullible in the face of sophisticated claims. Even as experts like Ekman himself condemn the manner in which commercial emotion recognition technologies are currently marketed, the market is steadily growing.[38]

Secondly, the ready supply of camera and surveillance infrastructure put in place due to the proliferation of facial recognition technologies means that emotion recognition systems are simply another layer of surveillance to be

34 Matt Simon, 'Fantastically Wrong: The Silly Theory That Almost Kept Darwin From Going on His Famous Voyage' *Wired* (21 January 2015) <https://www.wired.com/2015/01/fantastically-wrong-physiognomy/>.

35 Alexander Todorov (n 33).

36 Blaise Aguera Y Arcas, Margaret Mitchell, and Alexander Todorov, 'Physiognomy's New Clothes' (7 May 2017) <https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a>.

37 Javier Sanchez Monedero and Lina Dencik (n 11); Vidushi Marda and Shazeda Ahmed (n 5).

38 Madhumita Murgia, 'Emotion Recognition: Can AI Detect Human Feelings from a Face?' *Financial Times* (11 May 2021) <https://www.ft.com/content/c0b03d1d-f72f-48a8-b342-b4a9261094 52>.

plugged into existing networks. And finally, the use of emotion recognition is often opaque and invisible, making it difficult for civil society and academia to uncover in the absence of explicit mentions of the use of these technologies by the authorities themselves.

Given research on emotional expressions being culturally specific, there may be a temptation to collect data on Indian expressions and Indian faces to achieve greater accuracy, akin to experiences elsewhere.[39] It is crucial to note here, that the findings in this section do not lead to that being a viable next step - it is both dangerous and ill-conceived. Firstly, accuracy is a non-starter with respect to emotion recognition systems. The assumption that emotions can be inferred at all is scientifically dubious. Embarking on efforts to "accurately" do so not only actively ignore scientific evidence, it also obfuscates the fundamental issues of human rights and human dignity at play. Secondly, this will become an exercise in wide scale data collection for a vague, open-ended reason, cloaked under a "good" use case. It is also important to bear in mind that India is home to a multitude of cultures, and over 2000 ethnic groups - debiasing datasets in general has significant limitations, but attempting to do so in the context of emotion recognition is pursuing a fundamentally flawed premise.[40]

2. Data Governance Approach

A popular response to the dangers arising from the use of biometric technologies has been to put in place robust data protection safeguards to regulate and mitigate harms that arise out of them. Data protection is a crucial regulatory tool that can solve for myriad harms and risks arising from emerging technologies. In the case of emotion recognition technologies, however, data protection is an insufficient regulatory tool.

Traditional data governance approaches for the collection, processing, sharing and flow of data will amount to little to no protection for individuals subject to these technologies given the inherently problematic foundations on which emotion recognition technology is built. Data protection legislation may also become instrumental in legitimizing the use of these technologies through the illusion of procedural or even substantive guardrails. But the problem with emotion recognition is not only that it collects and processes sensitive personal data - it is that it exists at all. These technologies significantly violate individual dignity, fundamental rights, and impact individual and communities' access to opportunities, all while these systems cannot do what they purport to. They are intrinsically oppressive - as Adrian Daub has stated before, physiognomic logics such as the ones embedded in emotion

39 Vidushi Marda and Shazeda Ahmed (n 5).

40 Vidushi Marda, 'AI Bias Is Not Just a Data Problem' *Forbes India* (10 August 2021) <https://www.forbesindia.com/article/ai-special/column-ai-bias-is-not-just-a-data-problem/69693/1>.

recognition systems are not dangerous because they work, but rather because people *believe* that they do.[41]

The implicit assumptions of data protection law: Data protection law is traditionally placed to ensure and enable *fair and secure processing of data*, not necessarily to prevent it. It assumes that data collection, sharing and process will happen; that technologies will be built; and that their use should be regulated and facilitated through safeguards and institutional mechanisms respectively.

India's most recent attempt towards a data protection law was put in motion in 2019, and withdrawn in 2022. During hearings in *K.S Puttaswamy v. Union of India*, in which the Supreme Court affirmed the right to privacy in India, the Government declared its intention of setting up a Committee of Experts to deliberate on a data protection framework for India. The Committee, in its own words, was to facilitate the Government's vision to *"to unlock the data economy, while keeping data of citizens secure and protected."* [42]Discussing its vision for personal data collection, use and sharing in the digital economy, the Committee envisions *"a polity where the individual is autonomously deciding what to do with her personal data, entities are responsibly sharing such data and everyone is using data, which has immense potential for empowerment, in a manner that promotes overall welfare."* [43]

In other words, the existence of technologies is not only taken for granted, but also facilitated through data protection frameworks. In India, the draft Data Protection Bill 2021 contemplated creating "*a framework for organisational and technical measures in processing of data*" and laying down norms for cross border transfers.[44] This is similar to data protection efforts elsewhere. The EU's GDPR for instance, is meant to *"ensure the free flow of personal data within the Union and the transfer to third countries and international organisations"*[45] In 2021, one of several amendments made to the draft Data Protection Bill in India included the words "*to ensure the interest and security of the State"* in the Preamble, further cementing the fact that ideals of national security may and will come to bear on the protection of personal data.[46] While the draft Bill has since been withdrawn, reports at the time of writing suggest that a new bill is in the pipeline.

---

41 Adrian Daub, 'The Return of the Face' (*Longreads*, 3 October 2018) <https://longreads.com/2018/10/03/the-return-of-the-face/>.

42 Committee of Experts under the Chairmanship of Justice B. N. Srikrishna, 'A Free and Fair Digital Economy' (Ministry of Electronics and Information Technology) <https://www.meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf>.

43 ibid 9.

44 The draft Data Protection Bill, 2021(11 September 2019) ("DPB 2021"), preamble <https://prsindia.org/files/bills_acts/bills_parliament/2019/Joint_Committee_on_the_Personal_Data_Protection_Bill_2019.pdf>.

45 Regulation (EU) 2016/679 of the European Parliament and of the Council of 7 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] L119/1, recital 6.

46 See Report of the Joint Committee on the Personal Data Protection Bill, 2019 (16 December 2021), <http://loksabhaph.nic.in/Committee/CommitteeInformation.aspx?comm_code=73&tab=1&gt>

The urgent need for banning emotion recognition technologies is thus ill-placed under a data protection regime, except in the instance where a blanket ban or red line is drawn with respect to the design, development, deployment and use of these technologies. This requirement applies across all kinds of regulation. Under the EU AI Act, for instance, emotion recognition has been classified as posing limited risks to individuals generally, requiring only transparency when individuals are subject to such systems.[47] In the case of law enforcement use cases, emotion recognition may also be considered as posing high risk, and consequently needing to comply with a number of requirements under the Act. At no point, however, do emotion recognition systems get classified as posing unacceptable risks, and therefore are not banned under the EU AI Act - failing to address the harms that these technologies inflict on individuals and communities.[48]

A striking feature of data protection is that *data* of citizens is kept at the centre of its focus, as opposed to the citizens themselves. There is little to no space to question the impact of technologies on individuals - as the lion's share of efforts are geared towards ensuring that processing is conceivably fair and safe.[49]

This is particularly important to consider in the context of surveillance systems. Privacy law expert Graham Greenleaf has argued that the emphasis on fair and safe processing alone is not enough as it, "*obscures the broader issues of the extent of surveillance that a democratic society should accept. What degree of surveillance is too intrusive, unforgiving or dangerous, irrespective of how fairly and openly it is done*?".[50] Greenleaf further suggests that privacy principles or laws should instead be measured by "*the capacity they have to place limits on the extent of surveillance carried out, and, where appropriate, to stop proposed surveillance altogether*".[51]

Balancing competing interests: Emotion recognition applications are increasingly marketed, for the purpose of public sector use, under the umbrella of "good" use cases - to ensure safety and security of women, to keep public areas safe, to detect deceptive individuals and infer the truth of their inner states, and for the purpose of private sector use, under the umbrella of "efficient" use cases - to identify best candidates for a job, to make children better students and to keep drivers and passengers safe.

---

47 See Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts, art 3(34), art 52 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>.

48 'Europe: Artificial Intelligence Act Must Protect Freedom of Expression and Privacy' (*ARTICLE 19*) <https://www.article19.org/resources/europe-artificial-intelligence-act-must-protect-freedom-of-expression-and-privacy/>.

49 Woodrow Hartzog and Neil Richards, 'Privacy's Constitutional Moment and the Limits of Data Protection' (2021) 61 Boston College Law Review 1725.

50 Graham Greenleaf, 'Stopping Surveillance: Beyond "efficiency" and the OECD' (1996) 3 Privacy Law & Policy Reporter (Prospect Publishing).

51 ibid; Also see Neil M. Richards, 'The Dangers of Surveillance' [2013] Harvard Law Review 1934.

Under the recently withdrawn Indian data protection bill, the Government could choose to exempt any agency from provisions of the entire Act, if it is satisfied, *inter alia,* that it is necessary or expedient to do so, *"in the interest of sovereignty and integrity of India, the security of the State, friendly relations with foreign States or public order*". These exemptions were not only widely construed (public order is notoriously broad to interpret); they are also to be determined against the standards of necessity *or* expediency (the latter of which is not legally defined), providing the State with wide powers to exempt itself from provisions of the Act through executive orders that will not be published in the public domain. These envisioned uses fall under the categories of "national security" or "public order", among others, most of which enjoy wide exemptions under data protection law.[52] This means that even the limited protections like data minimisation and purpose limitation are absent in this context as in reality, they will not come to bear on emotion recognition technologies. While balancing provisions against business and state interests, and individual rights, recent trends in data protection legislation will more often than not *enable* data collection, as opposed to precluding it.[53]

The exemptions to restrictions and safeguards contemplated under the final version of the Act also became broader. The Srikrishna Committee in its 2018 report stated, "*The data protection law will enable an exemption to the processing of personal or sensitive personal data if it is proportionate and necessary in the interest of the security of the state and is pursuant to a law that meets the test of constitutionality. Further, any restriction on privacy must be proportionate and narrowly tailored to the stated purpose.*"[54] In the final version available in the public domain, this clause had significantly changed from exemptions being pursuant to a law and depending on the "security of the state", to being valid even if passed by executive order if considered necessary or expedient by the State, by a just, fair and proportionate procedure.[55]

India's current efforts towards a national Automated Facial Recognition System, indicate that the national security and public order exceptions will be enforced adversely to individual rights.[56] Beyond India, similar trends occur. Facial recognition is deployed for the purposes of public security, national security and public order in

---

52 This is the case globally. For instance, The GDPR also recognises that personal data protection is not absolute, and must be balanced with fundamental rights, in accordance with the principle of proportionality. See: General Data Protection Regulation (n 45) recital 4. Margot E Kaminski, 'Binary Governance: Lessons from GDPR's Approach to Algorithmic Accountability' (2019) 92 Southern California Law Review 1529.

53 Woodrow Hartzog and Neil Richards (n 49).

54 Committee of Experts under the Chairmanship of Justice B.N., 'A Free and Fair Digital Economy: Protecting Privacy, Empowering Indians' (2018) 128 <https://www.meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf&gt>.

55 DPB 2021 (n 44) clause 35.

56 Vidushi Marda, 'Every Move You Make' *India Today* (29 November 2019); Vrinda Bhandari, 'Facial Recognition: Why We Should All Worry about the Use of Big Tech for Law Enforcement' <https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/the-future-of-democracy-in-the-shadow-of-big-and-emerging-tech-ccg-248.pdf>; Smriti Parsheera, 'Adoption and Regulation of Facial Recognition Technologies in India: Why and Why Not?' (Data Governance Network 2020) <https://datagovernance.org/report/adoption-and-regulation-of-facial-recognition-technologies-in-india>

the US, UK, China, Brazil and beyond.[57] Facial recognition technology has also been approved in Denmark to enforce private ban lists at a football stadium given the substantial public interest allowed.[58]

This is not to say that a proportionality analysis with respect to face and/or emotion recognition will *never* result in protection of individual rights, but rather that this is more of the exception. In Marseille, France, for instance, the Administrative Tribunal disallowed the use of facial recognition in schools, given that it did not meet the tests of necessity and proportionality.[59] Sweden's Data Protection Authority levied its first fine under the GDPR on a local authority for trialing facial recognition on students, stating that there were less intrusive ways of tracking attendance.[60]

Even so, given emerging reports of State-led uses of emotion recognition technologies, and judging from the trajectory of this market world wide, it would appear that exemptions are afforded in precisely the relationships that need protection. This is not unique to India, as the limits of data protection as a framework are being seen in other jurisdictions as well. The GDPR asserts its scope at the outside, stating "*This Regulation does not apply to issues of protection of fundamental rights and freedoms or the free flow of personal data related to activities which fall outside the scope of Union law, such as activities concerning national security*".[61]

Narratives around security extend across use cases too, putting into focus the limited impact of purpose limitation clauses. For instance, face recognition in India was suggested on a trial basis by the Delhi High Court in 2018 to find missing children, and by the end of 2019, it was also used to track protestors exercising constitutional rights.[62] Even though the purpose is significantly different, the broad classification of the security of state, or public order enables overreach.[63] Further, because the use

---

57 Kashmir Hill, 'Wrongfully Accused by an Algorithm' *The New York Times* (3 August 2020) <https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html>; Samuel Woodhams, 'London Is Buying up Heaps of Facial Recognition Tech' *Wired* (27 September 2021) <https://www.wired.co.uk/article/met-police-facial-recognition-new>; 'One Month, 500,000 Scans: How China Is Using AI to Profile a Minority' *The New York Times* (14 April 2019) <https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html>; Charlotte Peet, 'Brazil's Embrace of Facial Recognition Worries Black Communities' *Rest of World* (21 October 2021) <https://restofworld.org/2021/brazil-facial-recognition-surveillance-black-communities/>.

58 Jesper Lund, 'Danish DPA Approves Automated Facial Recognition' *EDRi* (19 June 2019) <https://edri.org/our-work/danish-dpa-approves-automated-facial-recognition/>.

59 'When Bodies Become Data: Biometric Technologies and Freedom of Expression",' (ARTICLE 19 2021) <https://www.article19.org/wp-content/uploads/2021/05/Biometric-Report-P3-min.pdf>.

60 'Facial Recognition: School ID Checks Lead to GDPR Fine' *BBC* (27 August 2019) <https://www.bbc.com/news/technology-49489154>.

61 General Data Protection Regulation (n 41) recital 16.

62 'Delhi Police Facial Recognition Has Only 2% Accuracy: HC Told' *Business Standard* (23 August 2019) <https://www.business-standard.com/article/pti-stories/delhi-police-facial-recognition-software-has-only-2-per-cent-accuracy-hc-told-118082301289_1.html>; 'Upgrade Face Recognition Software: Delhi High Court' *Times of India* (24 August 2019) <https://timesofindia.indiatimes.com/city/delhi/upgrade-face-recognition-software-delhi-high-court/articleshow/70813797.cms>; Jay Mazoomdaar, 'Delhi Police Films Protests, Run Its Images through Facial Recognition Software to Screen Crowd' *The Indian Express* (28 December 2019) <https://indianexpress.com/article/india/police-film-protests-run-its-images-through-face-recognition-software-to-screen-crowd-6188246/>.

63 *Rise of COVID Surveillance and Facial Recognition* (Directed by Internet Freedom Foundation) <https://www.youtube.com/watch?v=I3eX5VGvloM&t=2750s>.

cases, particularly in the case of public sector applications are "good" and conceived to be in the public interest generally, the deployment of emotion recognition (and face recognition) technologies bypasses adequate scrutiny as companies offer these technologies free of cost, on a "trial" or "pilot" basis.[64]

Data protection is isolated from power and social realities: Data protection does not account for existing power structures, even as privacy is fundamentally concerned with power.[65] It aims to facilitate fair processing and safe storage and transfer of data. It may even centre individual autonomy, as the GDPR and India's data governance frameworks seem to suggest. But, data protection inherently does not question or challenge the existing status *quo* of power differentials and asymmetries. Protections like data minimisation, purpose limitation, notice and choice do not explore the world of justifying technologies and their uses but rather claim to secure and protect within it. Particularly in the case where the state is the data fiduciary, it also does not acknowledge existing dynamics between police institutions and historically marginalised communities, or structurally oppressed groups, or the impact of processing on such individuals. As a consequence, data protection is not the ideal place to deal with the dangers of surveillance, oppression, marginalization and criminalization of communities. It is chiefly concerned with efficient and safe data processing, instead of challenging the growth and ubiquity of surveillance.

Data protection frameworks treat individuals equally, i.e. as long as data is processed fairly, under the contours of the particular legislation, individuals have the same rights. However, in the context of face and emotion recognition and even predictive policing, research is increasingly demonstrating that the societal impact of technologies is layered, disproportionately presenting adverse outcomes from individuals of historically marginalised communities along the fault lines of income, religion, caste, among others. For instance, predictive policing in Delhi has been demonstrated to have a disproportionate impact on lower income and migrant colonies,[66] face recognition across the world is well-acknowledged to have a racial and gender bias,[67] and emotion recognition has been demonstrated to have a significant racial bias as well.[68] A blanket approach to regulating data processing with respect to these technologies ignores crucial social realities even as such systems and data controllers shape these societies.

---

64 Yuan Stevens and Ana Brandusescu, 'Weak Privacy, Weak Procurement: The State of Facial Recognition in Canada' (Centre for Media, Technology and Democracy 2021) <https://www.mediatechdemocracy.com/work/weak-privacy-weak-procurement-the-state-of-facial-recognition-in-canada>.

65 Daniel J. Solove, 'Privacy and Power: Computer Databases and Metaphors for Information Privacy' (2001) 53 Stanford Law Review 1393.

66 Vidushi Marda and Shivangi Narayan, 'Data in New Delhi's Predictive Policing System' [2020] FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency <https://dl.acm.org/doi/abs/10.1145/3351095.3372865>.

67 'Gender Shades' <http://gendershades.org/>.

68 Lauren Rhue, 'Racial Influence on Automated Perceptions of Emotions' [2018] SSRN Electronic Journal <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3281765>.

On the illusion of procedural safeguards: The procedural safeguards put in place through data protection regulations are also suspect in the context of emotion recognition. Not only are they ineffective in mitigating the fundamental risks posed by emotion recognition technology, they provide the illusion of fairness and safeguards without addressing root causes, at the same time acting as a legitimizing force for technology and technical infrastructures to be embedded within societies. Legal scholar Ari Ezra Waldman terms this phenomenon of reliance on procedural data protection safeguards as a stand in for real compliance with privacy law "legal endogeneity".[69] Compliance with data protection is increasingly recognised as a way of reducing corporate risk and responsibility, and decreasingly so as a real way of protecting privacy *of the individuals themselves*. Procedural safeguards as an end in and of itself simply buys time for problematic technologies like emotion recognition to proliferate and be firmly embedded in societies, affording this luxury under the wide umbrella of "innovation" and "modernization" at the cost of fundamental rights and individual protections. As Stark and Hutson have argued, "*The issue of physiognomy is not one of implementation; the morality of physiognomy is not resolvable through changes to input data and deployment*".[70]

Conclusion

Barring an explicit, blanket and water-tight rejection of emotion recognition technologies for their inherently invasive, discriminatory and problematic nature, data protection is a toothless and insufficient regulatory mechanism in context of emotion recognition technologies in particular and physiognomic AI more generally.

Beyond data protection law, a number of regulatory tools and levers should be used in the context of emotion recognition technologies. Constitutional challenges, for instance, are one way to secure overarching bans on the use of emotion recognition technologies given the direct impact on human dignity and autonomy, and subsequently on fundamental rights like privacy, freedom of expression, freedom of assembly, non-discrimination and the right against self-incrimination.

Another approach could be to challenge private power and corporate control over which technologies are built, how they are tested, and when they are deployed, through a closer look at procurement processes, or even using existing antitrust provisions could be effective. At present, civil society and academia is only invited to the table once all the building blocks are in place, or when a pilot or tender is being publicized. Challenging private actor's power and discretion must start earlier in the process.

---

69 Ari Ezra Waldman, 'Privacy Law's False Promise' (2020) 97 Washington University Law Review <https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=6386&context=law_lawreview>.

70 Luke Stark and Jevan Hutson (n 6).

A big reason as to why emotion recognition technologies are beginning to proliferate around the world is because institutional appetite - whether private or public institutions - for the procurement and use of surveillance technology is heightened. The willingness to experiment or trial new technologies, regardless of their impact on fundamental rights, highlights the need for adequate safeguards at the time of procurement and assessment as well.

Finally, reckoning with emotion recognition systems as a sociotechnical system, for state actors, companies, academia, civil society and the public at large is a foundational step towards challenging their use in societies. The issues at play given the use of emotion recognition are nefarious, structural and fundamental, and require a significant overhaul of existing assumptions and institutional practices. The need of the hour is to ban, not optimise for, emotion recognition systems.

Searching for a Room of One's Own in Cyberspace: Datafication and the Global Feministation of Privacy

Anja Kovacs[1]

Introduction

In May 2017, in the context of a Supreme Court case questioning the constitutionality of India's Aadhaar or unique ID program, the country's then Attorney-General, Mukul Rohatgi, made some controversial claims. When opponents attempted to counter the collection of biometric data by linking privacy and bodily integrity, Mr. Rohatgi labelled their arguments 'bogus', and added that Indians' right to their body was in any case not absolute.[2] Outrage ensued on social media and in newspaper columns about the latter point in particular.[3] But while it was disappointing to see the Attorney-General restating this principle in court, women and gender and sexual minorities knew that his words reflected reality. Whether in social life or in law and jurisprudence, their right to their bodies and related privacy sometimes is not recognised at all.

Over the past few years, growing attention has been paid to the ways in which gender and sexuality intersect with privacy concerns in the digital age. Whether through social, corporate, or state surveillance, such work highlights that women and sexual and gender minorities are at particular risk when a loss of privacy occurs as a consequence of digitisation and datafication.[4] But what the above anecdote draws attention to is that, if we are to examine how privacy and gender intersect, instances in which there is a loss of privacy should not be the only focus of our attention.

---

1 Anja is the Founder of Feminist Futures as well as a Non-Resident CyberBRICS Fellow at the Fundação Getulio Vargas (FGV) in Rio de Janeiro, Brazil. She can be reached at anjakovacs@gmail.com.

2 Amit Anand Choudhary, 'Citizens Don't Have Absolute Right over Their Bodies: Government' *Times of India* (3 May 2017) <https://timesofindia.indiatimes.com/india/citizens-dont-have-absolute-right-over-their-bodies-government/articleshow/58486260.cms>; DownToEarth Staff, 'Who Has Rights over a Citizen's Body? New Twist in Aadhaar Controversy' *DownToEarth* (3 May 2017) <https://www.downtoearth.org.in/news/governance/who-has-rights-over-citizens-body-new-twist-in-aadhaar-controversy-57754>.

3 'Aadhaar Case: Mukul Rohatgi Is Wrong. "Bodily Integrity" Is Sacrosanct' *Hindustan Times* (5 May 2017) <https://www.hindustantimes.com/editorials/aadhaar-case-mukul-rohatgi-is-wrong-bodily-integrity-is-sacrosanct/story-EghyEtXCUDkaw3RQ9TJ9qO.htm>.

4 Tatiana Dias and others, 'Mother in a Click: Pregnancy as a Jackpot for the Datasucker' (*Chupadados*, 2014) <https://chupadados.codingrights.org/en/vc-e-oq-vc-clica/>; Anja Kovacs, '"Chupke, Chupke": Going Behind the Mobile Phone Bans in North India' [2017] Gendering Surveillance <https://genderingsurveillance.internetdemocracy.in/phone_ban/>; Nayantara Ranganathan, 'A Handy Guide to Decide How Safe That Safety App Will Really Keep You' <https://genderingsurveillance.internetdemocracy.in/safety-app/>; Nayantara Ranganathan, 'Caution! Women at Work: Surveillance in Garment Factories' <https://genderingsurveillance.internetdemocracy.in/cctv/>; Vanessa Rizk and Dahlia Othman, 'Quantifying Fertility and Reproduction Through Mobile Apps: A Critical Overview' (2016) 22 Arrow for Change 13; Nicole Shephard, 'Big Data and Sexual Surveillance'.

In addition, as I will further explore below, the construction of privacy *itself* has been deeply gendered, as women and gender and sexual minorities are often at the receiving end of forms of privacy that are subordinating, rather than equalising.[5] Instead of enabling greater freedom, privacy then becomes a duty, a responsibility, the maintenance of which a woman or person belonging to a gender or sexual minority can and is being held accountable for. Privacy becomes something that is intended to keep their world small and restricted, rather than enabling exploration and expansion.

It is my contention in this essay that, as a result of pervasive datafication, we are now witnessing a generalisation of such problematic interpretations of privacy, to include and affect *everybody*. Although datafication is fundamentally reconfiguring our bodies and our lives,[6] a comprehensive rethink of what it means to substantially protect privacy in this context remains lacking. The result is that the watered-down, inferior version of privacy that women and sexual and gender minorities historically have been faced with is now extended to all. Those who are more privileged will continue to be less (negatively) affected than those who are marginalised, in one or more ways. But nobody can escape completely. We are effectively witnessing a *global feminisation of privacy*.

In what follows, I will first examine, in part one, in what ways dominant understandings of privacy have been gendered and how such gendering has been reflected in Indian jurisprudence. I will outline how privacy has been mobilised and interpreted in ways that have entailed a fundamental curtailment of the decisional autonomy of women and gender and sexual minorities, and of their ability to engage in self-determination. In part two, I will then argue that, in the age of datafication, this predicament now presents itself to all of us, as a result of three trends in particular: the specific ways in which consent and anonymity are mobilised by surveillance capitalism (and government) as key tools to drive the datafication of our lives; the resulting reconfiguration of the public and the private; and the portrayal and treatment of data as by default disembodied and deterritorialised. It is these three trends that lie at the heart of the global feminisation of privacy.

1. The Gender of Privacy

Privacy as a right has always had a somewhat ambivalent standing in feminism.[7] Revisiting the dominant understandings of the concept of privacy can explain why this is so. It will also allow me to establish the first part of my thesis: i.e. that privacy is gendered, or to be more precise, that the privacy protections that women, as well as

5 Anita L Allen and Erin Mack, 'How Privacy Got Its Gender' (1990) 10 Northern Illinois University Law Review 441.

6 Anja Kovacs, 'When Our Bodies Become Data, Where Does That Leave Us' (*Deep Dives*, 28 May 2020) <https://deepdives.in/when-our-bodies-become-data-where-does-that-leave-us-906674f6a969>.

7 Anita L Allen, 'Privacy' in Alison M Jaggar and Iris M Young (eds), *A Companion to Feminist Philosophy* (Blackwell 1998).

sexual and gender and minorities, have traditionally enjoyed are not only not the same as those that men have, but also that they are not unequivocally a positive good.

1.1. Privacy and the Home

Historically, and perhaps even today, dominant understandings of privacy have focused on the home, the domestic, as the locus of a set of both spatial and relational or socio-institutional conceptions of privacy. As Samuel Warren and Louis Brandeis put it in their seminal 1890 essay on the right to privacy, a 'man's home is his castle', and this man, not his government, is its ruler.[8] The home thus emerged as a space of exclusion, where one can exercise the right to be alone - or at least to be left alone by the state.

1.1.1. The home as a black box

Such constructs of privacy have for long been deeply criticised by feminists. By making the home, not the individual, the basic unit for privacy, they have argued, these notions disregard the unequal power relationships that exist within the household, at the expense of those more vulnerable in the equation.[9] With the protection of the home as a private space also came the designation of a whole range of relations and activities centred around the domestic sphere, such as the family and marriage, as private[10] and thus, to be excluded from state intervention. Moreover, the shape such relations and institutions would take often would primarily benefit the interests of the men of the household. As homes, families, and marriages were designated men's castles to rule, 'the private' thus often functioned as a flag to cover up the oppression of, and violence against, women that takes place in homes.

When the Indian Supreme Court, in the absence of a specific provision guaranteeing this right, for the first time read a right to privacy into the Indian Constitution in 1975,[11] the judgment reflected such an intertwining of spatial and socio-institutional conceptions of privacy driven by patriarchal, heteronormative ideals. It said: 'Any right to privacy must encompass and protect the personal intimacies of the home, the family, marriage, motherhood, procreation and child rearing'. Motherhood, but not fatherhood, as Gautam Bhatia has pointed out.[12]

---

8 Samuel D Warren and Louis D Brandeis, 'The Right to Privacy' (1890) 4 193.

9 Catharine A MacKinnon, *Feminism Unmodified: Discourses on Life and Law* (1987); Linda J Nicholson, *Gender and History: The Limits of Social Theory in the Age of the Family* (1986); Frances Olsen, 'Constitutional Law: Feminist Critiques of the Public/Private Distinction' (1993) 10 319; Elizabeth M Schneider, 'The Violence of Privacy' (1991) 23 Connecticut Law Review 973; Ruth Gavinson, 'Feminism and the Public/Private Distinction' (1992) 45 Standford Law Review 45.

10 Gautam Bhatia, 'The Constitution and the Public/Private Divide: T. Sareetha vs. Venkatasubbaiah' <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3010972>.

11 *Gobind v. State of MP* (1975) 2 SCC 148.

12 Bhatia (n 10).

Because indeed, the protection of non-interference foreseen was principally for the benefit of the institutions concerned, not for individual rights or intimacy as such. For example, as Aparna Chandra (2017) has noted, while it is often held that sexual relations are a private matter, in many ways it is not sex but 'marital sex' that is protected by privacy.[13] It is for this reason that, for example, section 9 of the Hindu Marriage Act, 1955 could for many years coexist with the criminalisation of consensual homosexual intercourse under section 377 of the Indian Penal Code. Section 9 of the Hindu Marriage Act allows for the reconstitution of conjugal rights where a spouse has deserted the other 'without reasonable excuse' and is often seen as a tool for women's oppression.[14] Only in 2018 was the criminalisation of consensual homosexual intercourse finally struck down by the Supreme Court.[15]

To its credit, the *Puttaswamy* judgement of 2017, the landmark privacy judgement pronounced by India's Supreme Court,[16] seemed to break with this tradition. As I will discuss in detail below, it acknowledged these feminist critiques and highlighted the importance of privacy for individual autonomy and decision-making. Yet even this judgement was not free from this tension. For example, in his opinion, Justice Bobde referred, among other things, to the 'well-established rule in the Ramayana' that 'a woman ought not to be seen by a male stranger' as evidence that a 'well-developed sense of privacy' existed 'even in the ancient and religious texts of India'.[17] Similarly, he noted, 'Religious and social customs affirming privacy also find acknowledgement in our laws, for example, in the Civil Procedure Code's exemption[18] of a pardanashin's lady's appearance in Court'.[19] While the latter provision may have its uses, it is noteworthy that Justice Bobde never asked whose privacy is really sought to be protected in these instances: the woman's, or her husband's and his family's?

This reminds us not only that the private sphere is never isolated from government influence; in addition, heteronormative, patriarchal notions of gender often shape government imagination of what is private and what is not

---

13 Aparna Chandra, 'Privacy and Women's Rights' (2017) 52 Economic and Political Weekly; Martha C Nussbaum, 'Is Privacy Bad for Women' (*Boston Review*, 21 July 2014) <https://bostonreview.net/world/martha-c-nussbaum-privacy-bad-women>.

14 The constitutionality of section 9 of the Hindu Marriage Act, too, is currently being challenged in the Supreme Court. The arguments revolve to an important extent around the patriarchal nature and impact of the section, its gender neutral language withstanding. See Samanwaya Rautray, 'SC to Examine Whether Forcing Woman to Stay with Husband against Her Will Is Violative of Her Rights' *Economic Times* (5 March 2019) <https://economictimes.indiatimes.com/news/politics-and-nation/sc-to-examine-whether-forcing-woman-to-stay-with-husband-against-her-will-is-violative-of-her-rights/articleshow/68274137.cms>.

15 *Navtej Singh Johar And Ors. v. Union of India* (2018) 10 SCC 1.

16 *Justice K.S.Puttaswamy (Retd). v. Union of India And Ors.* (2017) 10 SCC 1.

17 ibid para 23.

18 Section 132 of the Code of Civil Procedure, 1908.

19 *Puttaswamy* (n 16) para 23.

as much as society's. And to the extent that this is true, such notions tend to be reflected in law as well.[20]

1.1.2. Privacy and space

Spatial notions of privacy that are focused on the home also disregard the fact that for many, including many women, people belonging to gender and sexual minorities, and poor people, the home in fact provides few opportunities for solitude, or to be left alone - sometimes simply because of space constraints. To be fair, not all women consider an absence of solitude problematic;[21] they may well consider companionship and care more important goals, and things to be cherished.[22] But as one of the main loci of socialisation for women in particular, the home can also become a deeply stifling space,[23] especially when they attempt to develop their individual identity in ways counter to established norms.

Many in such situations turn, then, to the public sphere, rather than the private, to find privacy and autonomy. The young people sharing intimacies on public transport and in parks and cinema halls across India attest to this. For couples in same-sex, inter-faith, or inter-caste relations whose families disapprove, such opportunities may be especially important.

The public sphere is deeply gendered as well, of course: it requires women and gender and sexual minorities to always demonstrate a clear purpose for being in public and to thus confirm that their place really is in the private.[24] Such activities, therefore, are not without risk. But for those willing to take that risk, the anonymity that public space provides, especially in cities, can be a key tool in aiding the transformation of unequal gendered relations in the household.[25]

For those at the vulnerable end of unequal power relations, privacy in public can be as important as privacy at home.

1.2. Privacy and Autonomy

This brings us to a second key understanding of privacy, one that has gained prominence in recent years: privacy as a core element of autonomy and agency

20 Allen, 'Privacy' (n 7); Chandra (n 13).

21 Chandra (n 13).

22 Allen, 'Privacy' (n 7).

23 Chandra (n 13).

24 Shilpa Phadke, Sameera Khan and Shilpa Ranade, *Why Loiter? Women and Risk on Mumbai Streets* (Penguin 2011).

25 This point builds on Ambedkar's work which, addressing a slightly different context, urged India's dalits to move to the cities. The anonymity these provide, Ambedkar argued, would give them a much better chance at escaping many of the pressures of casteism, and thus to build better lives for themselves, than India's villages would. See Jesús Francisco Cháirez-Garza, 'Touching Space: Ambedkar on the Spatial Features of Untouchability' (2014) 22 Contemporary South Asia 37.

in decision-making, crucial for the development of our capacity for self-determination and, thus, for our subjectivity.[26]

Central to this aspect of privacy is our ability to engage in the management of our personal boundaries - whether physical or digital - as we see fit. Whenever we take a decision about what to reveal or not to reveal about ourselves to others, be it in social settings or more formal ones, we engage in such boundary management. These decisions are always contextual and dynamic: they change as our relationships change or as new situations emerge. In this way, boundary management allows us to create breathing room to validate our own experiences, beliefs, feelings and desires. Especially when these do not align with dominant norms, this is critical to living a life with dignity. Through boundary management, it becomes possible for us to be deeply social, relational beings, pervasively shaped by our social worlds, while at the same time being able to take a step back and develop a critical perspective on those worlds around us, and through this, our capacity for self-determination.[27]

If solitude may not matter that much to some women and people belonging to gender and sexual minorities, boundary management does. Yet the ability to decide for oneself how much to reveal and to whom, often remains denied to them, as others arrogate to themselves the right to take these decisions for them.

1.2.1. Gender, sexuality and autonomy in Indian jurisprudence

How has the Indian judiciary fared in this regard? Since at least 1983, Indian jurisprudence has considered the contradictions between laws that protect the institutions of family, marriage, motherhood and procreation, on the one hand, and Indian women's autonomy and agency with regard to these areas of life, on the other.[28] Yet, the judiciary's track record on this count since then has been uneven at best, and in all too many cases, it failed to recognise individual privacy and decisional autonomy where a challenge concerned the sphere of the family. Thus, allowing ambiguity to surround the status of notions of autonomy and personal decision.[29]

---

26 Julie E Cohen, 'What Privacy Is For' (2013) 126 Harvard Law Review.

27 ibid; Kovacs (n 6).

28 See, for example, the detailed analysis of *T. Sareetha v. T. Venkatasubbaiah*, AIR 1983 AP 356, by Bhatia. In this case, Justice PA Choudary of the Andra Pradesh High Court struck down section 9 of the Hindu Marriage Act as unconstitutional, on the grounds that it violates women's personal privacy, bodily integrity and individual dignity, as well as furthering inequality between men and women. Justice Choudary's verdict was overruled by the Supreme Court a year later, in *Saroj Rani v. Sudarshan Kumar Chadha*, AIR 1984 SC 1562. For further examples, see Surabhi Singh, 'The Puttaswamy Effect: Exploring the Right to Abortion in India' (*Centre for Communication Governance at, National Law University, Delhi*, September 2021) <https://nluwebsite.s3.ap-south-1.amazonaws.com/uploads/the-puttaswamy-effect-exploring-the-right-to-abortion-in-india-ccg-5.pdf>.

29 Singh, ibid.

This finally seemed to change in 2017, when India's Supreme Court reaffirmed in *Puttaswamy vs. Union of India* that the Indian people enjoy a fundamental right to privacy under their Constitution: privacy as autonomy and decision-making was one of the pillars around which the judgement was framed. Thus, this decision had the potential to open up a clear path to challenge any attempt, in law or jurisprudence, to undermine the decision-making agency and autonomy of women and sexual and gender minorities regarding even the most intimate aspects of life.[30] This was even more so because the importance of privacy for autonomous decision-making regarding one's body and to preserve bodily integrity and dignity, in particular, was discussed in the judgement at length. With this, the judgement implicitly recognised the centrality of the body, and the control over sexuality, in keeping inequalities intact - as well as in challenging them.

Yet challenges remained. As the judgement also reiterated the language of privacy centred on the home and household, whether in spatial or in relational or socio-institutional terms, it remained unclear how any conflicts between the individual rights of women and gender and sexual minorities, on the one hand, and the protection of institutions related to the home and family, on the other, would be resolved.[31] Would individual freedom actually prevail?

1.2.2. After Puttaswamy

It seems like something might finally be shifting in the law. In 2018 alone, for example, the Supreme Court delivered several landmark verdicts. It finally decriminalised homosexual relations between consenting adults;[32] reaffirmed the right of adult women to choose their own life partners and faith;[33] and decriminalised adultery.[34] Until then, the crime of adultery was deemed to be committed when a man slept with the wife of another man - as if women are men's property, and as if women can only be victims, not agents in this scenario (not to mention that the law didn't even recognise the possibility of same-sex relationships).

There is much to celebrate then. Yet constructions of 'home and marriage as sacred private spaces' remain alive in law.[35] For example, section 375 of the

30 Chandra (n 13); Anja Kovacs, 'How Privacy as a Fundamental Right Brings New Hope to India's Marginalised' *Hindustan Times* (30 August 2017) <https://www.hindustantimes.com/analysis/how-privacy-as-a-fundamental-right-brings-new-hope-to-india-s-marginalised/story-3hNuzNyUkK9LtYD8CjyNpI.html>.

31 Chandra (n 13).

32 *Navtej Singh Johar* (n. 15)

33 *Shafin Jahan v. Asokan K.M & Ors.*, AIR 2018 SC 357.

34 *Joseph Shine v. Union Of India*, WP (Crl.) 194/2017.

35 Chandra (n 13).

Indian Penal Code, which criminalises rape, continues to explicitly exempt marital rape. Forcing a woman to have sex without her consent in the context of marriage, although under challenge, remains legal in India for the time being.[36] In addition, albeit in varying degrees, across the judiciary courts continue to apply patriarchal norms in formulating their verdicts. For example, in 2021, the High Court of Haryana and Punjab dismissed a petition for protection filed by a couple in a live-in relationship who feared violence from their families, arguing that such relationships are neither 'morally nor socially accepted'.[37] This happened even though, as another bench of the same court pointed out in 2021 as well, nothing in the law forbids such relationships.[38] Even bills currently under discussion continue to contain provisions that undermine the decision-making agency of already marginalised people. For example, section 16 of the Trafficking in Persons (Prevention, Care and Rehabilitation) Bill, 2021,[39] allows for a Magistrate to send actual or perceived victims of trafficking to a rehabilitation home without having to even so much as ask the person for their opinion. Instead, the person in question has to make an application for their release, which the Magistrate can reject if he believes the application was not made voluntarily.

For this legacy to be completely undone, a lot more judicial and legal work is, thus, needed.

1.3. The Privacy Predicament

What women in India and, in varying degrees, around the world, continue to suffer from, then, is what Anita Allen has called 'the privacy predicament'.[40] They have too much of the '*wrong* kinds of privacy':[41] imposed modesty, chastity, reserve, and confinement, even isolation, in the 'privacy' of the home. They do not have enough privacy in the sense of adequate opportunities for 'replenishing solitude' or for boundary management, private choice, or autonomous decision-making.[42]

---

36 'Plea on Criminalisation of Marital Rape: Delhi HC Rejects Centre's Request for More Time, Reserves Verdict' *The Indian Express* (22 February 2022) <https://indianexpress.com/article/cities/delhi/marital-rape-criminalisation-pleas-delhi-hc-reserves-judgment-7783793/> accessed 22 April 2022.

37 *Gulza Kumari & Another v. State of Punjab & Ors.* CRWP No.4199 of 2021 (O&M).

38 *Pardeep Singh & Another v. State of Haryana & Ors.* CRWP-4521-2021 (O&M).

39 The Trafficking In Persons (Prevention, Care And Rehabilitation) Bill, 2021 <https://wcd.nic.in/sites/default/files/DRAFT%20TRAFFICKING%20IN%20PERSONS%20%28PREVENTION%2C%20CARE%20AND%20REHABILITATION%29%20BILL%202021%20%281%29.pdf>.

40 Anita L Allen, *Uneasy Access: Privacy for Women in a Free Society* (Totowa: Rowman & Littlefield 1988).

41 Anita L Allen, 'Gender and Privacy in Cyberspace' (2000) 52 Stanford Law Review 1175.

42 ibid 1179.

In these circumstances, privacy then is no longer a right, but a duty, a burden to be carried. And when their privacy is violated, it is women themselves who are held responsible. *She* should not have shared that picture. *She* should not have gone to that place.[43] Thus, as discussed before, even in the public sphere women have to demonstrate purpose to perform respectability and make evident that the private really is where they believe they belong.[44] Irrespective of how they behave, however, when something does go wrong, women themselves are generally held accountable 'for not being private enough', in ways that men simply are not.[45] And initiatives such as the deployment of facial recognition systems by law enforcement to detect 'women in distress' in public only reinforce notions that the public sphere really is a masculine domain.[46]

Unlike other human rights, privacy has been much more contradictory in its applications then. For privacy to do even half of the work we expect of it, the uneven way in which privacy is made accessible and the contradictory uses to which it is put need to be thoroughly interrogated. Without such an examination, we might end up inscribing into law a solution that is half-hearted at best and deeply damaging in the long term at worst. The experiences of many women and gender and sexual minorities in India who have attempted to mobilise the right to privacy to support their autonomous decision-making are instructive in this regard.

2. Privacy and Data

If women have always had to contend with the wrong kind of privacy and the associated burdens, this challenge now reasserts itself in the age of datafication — but no longer only for women. Rather, 'in their quest to make all of us increasingly legible, transparent, predictable, and manipulable, governments and private actors are fundamentally undermining our capacity to engage in the autonomous management of our bodies, selves, and lives as we see fit'.[47] Women, gender and sexual minorities and other marginalised groups will remain more vulnerable to these efforts and their impacts than more privileged sections of the population; yet nobody is excluded from the fundamental curtailment of our decisional autonomy and ability to engage in self-determination that data governance regimes at present entail.

This is what I mean by the global feminisation of privacy. Because, while some regions of the world may be better off because stronger regulation is in place, such

43 Allen, 'Gender and Privacy in Cyberspace' (n 41).

44 Phadke, Khan and Ranade (n 24).

45 Allen, 'Gender and Privacy in Cyberspace' (n 41).

46 Danya Hajjaji, 'Indian City Deploys Facial Recognition to Detect Harassed Women's Expressions' *Newsweek* (22 January 2021) <https://www.newsweek.com/indian-city-deploys-facial-recognition-detect-harassed-womens-expressions-1563761>.

47 Kovacs (n 6).

as Europe with its General Data Protection Regulation, this remains a worldwide phenomenon - albeit in varying degrees.

2.1. Key Tools Underpinning the Political Economy of Datafication_

Some aspects of the global changes that are underpinning this shift are, by now, well-known: intense datafication of our everyday lives; a securitisation of State-citizen relations that is in part driving this; and a shift to economic visions and policies that see surveillance capitalism as the driver of growth and well-being.[48]

Under the influence of these trends, two key tools have emerged that have contributed significantly to the global feminisation of privacy. The first is the widespread mobilisation of user 'consent' as a tool to legitimise contracts (between users and corporations or users and the state) that effectively undermine users' privacy. If we don't like how our data may be used, we simply shouldn't hand it over, i.e. we shouldn't consent, users are told. But such arguments leave out of consideration that users generally do not have 'the power to influence how this consent is defined, where it begins and ends, or what it looks like'[49] nor, in many cases, do they have the option not to consent. User consent in the digital age can therefore hardly be called meaningful. On the contrary, 'consent' here functions to effectively invisibilise, depoliticise, and even legitimise the new data infrastructures and deeply unequal power relations that shape them, while at the same time turning the protection of privacy into an individual responsibility.[50] As Lindsay Weinberg has noted, 'Privacy rights enacted through contracts largely protect the interests of corporations'.[51] Or, the state, as the case may be.

If the protection of their privacy for long has been a burden for women to carry as much as a right, current uses of consent tools ensure, in other words, that now all of us carry such a burden. Moreover, this challenge is increasingly heightened as consent tools are also used to legitimise the myriad practices - such as third party data sharing - that make boundary management by users effectively impossible.[52] In other words, the same tool that shoulders users with greater responsibility for the protection of their own privacy simultaneously curtails their ability to actually do so. If the capacity to engage in boundary

48 See e.g., Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Profile Books 2019).

49 Anja Kovacs and Tripti Jain, 'Informed Consent—Said Who? A Feminist Perspective on Principles of Consent in the Age of Embodied Data—A Policy Brief' (Internet Democracy Project 2021) <https://internetdemocracy.in/policy/informed-consent-said-who-a-feminist-perspective-on-principles-of-consent-in-the-age-of-embodied-data-a-policy-brief> accessed 26 October 2022.

50 ibid.

51 Lindsay Weinberg, 'Rethinking Privacy: A Feminist Approach to Privacy Rights after Snowden' (2017) 12 Westminster Papers in Communication and Culture 5.

52 Anja Kovacs and Tripti Jain (n 49); Kovacs (n 6).

management is crucial to preserve privacy as autonomy and decision-making, the way in which consent tools are deployed to legitimise datafication now undermine the ability of us all to do so effectively in the digital age.[53]

A second key tool to support the intense datafication of our lives further challenges our ability to effectively engage in boundary management: the practice of anonymising data. By anonymising data, 'institutions can argue they uphold the legal protections afforded to users in regard to individual privacy and concerns over discrimination'.[54] Yet even anonymous data can easily be mobilised to undermine a user's ability to engage in boundary management. In fact, that is often precisely its goal: in many cases, companies and states simply seek to figure out what 'type' of user you are, so as to slot you into different categories, based on which further action, targeting or excluding you, may or may not be taken. For example, algorithms used to decide who to advertise a particular job to need not know people's names; they rely instead on knowledge about a range of other attributes to take such decisions, including, in the case of one infamous algorithm developed by Amazon, people's gender.[55] While the persons affected may not even be aware that those deploying the algorithm possess this wealth of data about them, such exercises can nevertheless significantly affect the opportunities they get access to in life. If in the past, privacy has often been used as a flag to cover women's lack of autonomy within the household, it now 'conceals the non-sovereignty of online users who are governed through the commercialised capacity to distill patterns in aggregate data'.[56]

Whether offline or online, anonymity, and its contribution to enabling boundary management, has often been pivotal to the transformation of social relations. Yet in the age of datafication, meaningful anonymity is less and less available to us.

2.2. The Reconfiguration of the Public and the Private

There are, however, two additional shifts, underlying all these changes, that are crucial to understand why and how the global feminisation of privacy could have come about - and these have received far less attention so far.

---

53 Although consent will likely never be able to do all we are currently expected from it, a feminist analysis makes clear that existing consent regimes can be strengthened considerably. Among other things, this requires an acknowledgement that autonomy is always relational to be at the heart of any meaningful consent regime. For a detailed analysis of the kind of changes required and why, see Anja Kovacs and Tripti Jain (n 49).

54 Weinberg (n 51).

55 Miranda Bogen, 'All the Ways Hiring Algorithms Can Introduce Bias' (*Harvard Business Review*, 6 May 2019) <https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias>; Jeffrey Dastin, 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women' *Reuters* (11 October 2018) <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>.

56 Weinberg (no 51) 13.

The first crucial shift is the reconfiguration of the public/private divide. As I have outlined above, feminist political theory has for long critiqued the notion that the public and private are two separate, independent spheres of life. In the past, these critiques have primarily focused on bringing to light the gendered social relations that underlie the construction of both spheres and their interrelation. Now has been added to this the 'hybridisation of public and private life where the private is increasingly publicised, commodified, and subject to state and corporate surveillance'[57] — often by co-opting us into the surveillance of our own private lives.[58] Even the most intimate aspects of our existence are now made easily accessible to corporate and government interests.

Once again, the clean division between public and private life that liberal democratic theorists argued was essential to democracy is, thus, revealed as a mirage, and the private is effectively established as public. As Weinberg has pointed out, this time, however, this has largely happened without an accompanying liberatory politics such as that of the women's movement.[59] Privacy rights as currently conceptualised (and many who advocate for them) largely continue to maintain a dichotomy between public and private life. They do not recognise sufficiently the ways in which the political economy and cultural practices of datafication undermine the very dichotomy of the public and private, and the liberal idea of the sovereign subject, on which these rights continue to be largely based;[60] nor do they take into account that the subject is always relationally constituted. As a consequence, they once again fall short in protecting our rights, at the risk of becoming irrelevant. If we are to turn around the global feminisation of privacy, these realities need to be squarely faced and addressed.

2.3. The Myth of Disembodied Data

This brings me to a second important further shift that lies at the heart of the reconfiguration of the relation between the public and the private, and the rise of dataveillance that drives it: a new way in which the individual is believed to be constituted. No longer are we the 'juridical, rights-bearing subject of liberal democracy',[61] instead we are treated as what Deleuze has termed 'dividuals':[62] disembodied, deterritorialised beings, fragmented into masses of data points which can be aggregated and used for purposes of control by those who have

57 Weinberg (n 51) 8.

58 Kirstie Ball, Maria Laura Di Domenico and Daniel Nunan, 'Big Data Surveillance and the Body-Subject' (2016) 22 Body & Society 58.

59 Weinberg (n 51).

60 ibid.

61 ibid 7.

62 Gilles Deleuze, 'Postscript on the Societies of Control' (1992) 59 October.

both access to the data and the means to engage in such processes. Indeed, most of the value of datafication for corporations and governments lies in the analysis that aggregation allows for, even if our individual experiences of datafication may not make this self-evident.

If the notions of selfhood and rights, on which liberal democracy is based, presume an indivisible subject, how could such a shift happen in any rights-respecting society — and seemingly largely uncontested? The answer lies in how we understand what data is. Currently dominant understandings of data - whether furthered by governments, big tech companies, or start-ups - portray data as a layer that somehow penetrates everything, yet exists independently of the medium that has generated it.[63] Moreover, once generated, this data, just like a natural resource such as oil, is held to be simply 'out there', ready to be mined and used by companies and governments as they see fit.

As dividuals, we are treated as disembodied and deterritorialised because the argument is that data is disembodied and deterritorialised.

Yet such understandings of data often do not sit well with our experiences as users. After all, more and more, decisions that affect our physical bodies, their movements and actions are taken on the basis of our data bodies, and the claims our physical bodies may make, including to challenge such decisions, have less and less power in their own right. Rather than an independent layer or a disembodied mirror of our bodies, our experience, thus, tells us that our bodies and our data are closely intertwined - so much so that, as Irma van der Ploeg has pointed out, the line between our physical bodies and our virtual bodies really has become irrelevant.[64] Seeing that our bodies are always relationally constituted, the intense datafication of our lives necessitates, in other words, a veritable paradigm shift in the conceptualisation of our bodies: rather than treating the virtual merely as a reflection of the physical, our understanding of our bodies now needs to comprehensively incorporate both.[65] This also means that the broader data relations[66] in which we find ourselves, too, need to be centrally considered in any analysis of our embodied experiences and realities today, and of how these have come into being.

---

63 Katherine N Hayles, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics* (University of Chicago Press 1999).

64 Irma van der Ploeg, *The Body as Data in the Age of Information* (Kirstie Ball, Kevin Haggerty and David Lyon eds, Routledge 2012).

65 ibid.

66 'Data relations' refers to the range of relationships through which datafication is operationalised, including those that connect us to the economic and government actors who surveil us, or which datafication manifests in new digital forms, such as group based oppressions such as sexism and racism. Salomé Viljoen, 'Data Relations' (2021) 13 *Logic* <https://logicmag.io/distribution/data-relations/>. Also see Ulises A Mejias, *The Costs of Connection: How Data Is Colonising Human Life and Appropriating It for Capitalism* (Stanford University Press).

> Seeing that these connections are so rarely made, it is no surprise that the protection of our rights is in crisis. While bodily autonomy and integrity and autonomy over fundamental personal choices were also discussed at length in the *Puttaswamy* judgment, the impact of data governance regimes on these aspects, including through their impact on our capacity to effectively engage in boundary management under regimes of datafication, was not considered. They couldn't be, as the data-as-resource narrative has ensured that we, so far, simply lack the common vocabulary necessary to do so.[67] As long as the fact remains obscured that data and the impact of targeting people through data are always embodied, and thus always also have a material basis, data will continue to be treated as something that merely concerns 'informational privacy'. Would we have been so cavalier with how we understand consent or the value of anonymity online otherwise?

Way Forward

The global feminisation of privacy is then a direct consequence of these trends: in essence, it is the dissociation between bodies and data and the consequent removal of the embodied, relationally constituted realities of people and their lives from the data governance debate that has enabled the global feminisation of privacy in the age of datafication. The reverse is also true: if obscuring the entanglements of our physical bodies and our data is at the heart of the numerous shifts that have led to the global feminisation of privacy, bringing their deep interconnections, and the new context that shapes them, to light is a crucial first step in reversing this trend. In claiming that the link between data, privacy and bodily integrity is 'bogus', Mr. Rohatgi was simply incorrect.

In India, at least, it is an opportune time to do so. In the *Puttaswamy* judgement, we have finally found a firm legal basis to promote and protect the bodily privacy of women and gender and sexual minorities as a crucial element of their autonomy and decision-making, and a number of judicial decisions doing precisely that are already available. Recognising that in the datafied society, our bodies are fundamentally reconstituted to encompass not only flesh, blood, organs, emotions and senses, but also data, would now allow us to bring these long-overdue acknowledgments and their positive impact into the arena of data governance as well.

---

67 Anja Kovacs and Tripti Jain (n 49).

Notes on Contributors

Editors

- **Jhalak M. Kakkar** is Executive Director at the Centre for Communication Governance and a Visiting Professor at the National Law University Delhi. She works extensively on AI regulation and data governance, providing policy comments to the Indian government and international organisations like OHCHR on the design of AI regulation and data governance. Jhalak is an Expert member of the Global Partnership on AI (GPAI) Multistakeholder Experts Group Plenary and member of the Working Group on Data Governance and a member of the Board (Acadmic Alternate) of the Global Network Initiative. Prior to CCG, she has been a Visiting Researcher at Harvard Law School and a Programme Manager at PRS Legislative Research. Jhalak has an LLM from Harvard Law School (Fulbright-Nehru Masters Fellowship) and a law degree from the National University of Juridical Sciences, Kolkata.

- **Shashank Mohan** is a Programme Manager at the Centre for Communication Governance. His work is primarily focused on data protection, data governance, intermediary liability, and e-governance. Shashank is interested in studying the effects of digitization and the Internet on human rights, specifically the rights to privacy and free speech. He graduated from Symbiosis Law School, Pune in 2014.

- **Swati Punia** works with the Centre for Communication Governance on issues that lie at the intersection of technology, law and policy, and society. Her focus areas include privacy, data protection, data governance and emerging technologies. At present, she is examining the non-crypto blockchain ecosystem in India and studying its potential for addressing socio-economic challenges, creating inclusive e-governance models, and embedding privacy in the digital domain. Prior to joining CCG, Swati worked with a leading southern voice on fostering consumer sovereignty in digital economy. She is a lawyer by training and has earned certificates in digital trade and technology, cyber law, and corporate law.

- **Vrinda Bhandari** is a Rhodes Scholar, who graduated from the University of Oxford with a double Masters in Law and Public Policy, and received her undergraduate law degree from NLS Bangalore. She is a litigating lawyer in Delhi, and specialises in the field of digital rights, technology, and privacy. She has been a part of the Aadhaar challenge, the restoration of internet

in Jammu & Kashmir, the WhatsApp/Facebook dispute, Aarogya Setu, and the consultation process adopted for the National Digital Health ID. Most recently, Vrinda was involved in drafting LiveLaw's challenge to the Intermediary Rules, 2021 before the Kerala High Court. She has authored academic and policy research on information technology issues ranging from privacy data protection to intermediary liability.

Authors

- **Amber Sinha** works at the intersection of law, technology and society, and studies the impact of digital technologies on socio-political processes and structures. His research aims to further the discourse on regulatory practices around the internet, technology, and society. He is currently a Senior Fellow-Trustworthy AI at Mozilla Foundation studying models for algorithmic transparency. Amber was previously the Executive Director of the Centre for Internet and Society, India (CIS) where he led programmes on civil liberties research, including privacy, identity, AI, cybersecurity and free speech. His first book, *The Networked Public*, was released in 2019. He studied law and humanities at National Law School of India University, Bangalore.

- **Ameen Jauhar** is a Senior Resident Fellow at the Vidhi Centre for Legal Policy, leading its Centre for Applied Law & Tech Research (ALTR). At ALTR, he steers the organisation's engagements and independent research around Global South tech policy issues, with a focus on the Indian landscape. Ameen has written and advised policymakers and legislators on issues of data governance and artificial intelligence regulation. He also continues to work with the Supreme Court's E-Committee on the use of emerging technologies within the justice system.

- **Anja Kovacs** is a researcher, consultant, writer and public speaker. She is the Founder of Feminist Futures as well as a Non-Resident CyberBRICS Fellow at the Fundação Getulio Vargas (FGV) in Rio de Janeiro, Brazil. Prior to this, Anja was the Founder and Director of the Internet Democracy Project. Her research focuses on how to realise feminist visions of the digital in society, by exploring and addressing power imbalances in norms, governance and infrastructure. As a consultant on Internet issues, she has worked for, among others, the Independent Commission on Multilateralism, the United Nations Development Programme Asia Pacific and the UN Special Rapporteur on Freedom of Expression, Mr. Frank La Rue.

- **Arindrajit Basu** is currently a Non-Resident Research Fellow at the Centre for Internet & Society, India. He works primarily on the geopolitics and constitutionality of emerging technologies. Arindrajit is a lawyer by

training and holds a BA, LLB (Hons) degree from the National University of Juridical Sciences, Kolkata, and an LLM in public international law from the University of Cambridge, U.K.

- **Astha Kapoor** is the Co-founder of Aapti Institute, a Bangalore based research firm that works on the intersection of technology and society. She has over a decade of public policy and strategy consulting experience, with a focus on use of technology for welfare. At Aapti, Astha leads the Data Economy Lab, a vertical established to research and test new methods of data sharing, data stewardship and governance. Her recent work is focused on participative governance of data, and its use for building collaborative AI, through collective governance methods such as cooperatives. She serves on the advisory boards of the Data Trust Initiative (Cambridge University) and Indian Urban Data Exchange (IUDX).

- **Gangesh Varma** is a Senior Associate at Saraf and Partners focusing on technology and policy. Prior to joining the firm, he worked with prominent think tanks and international organisations, in the technology policy ecosystem. His areas of work range between internet governance and other several cutting-edge policy issues such as competition in the digital economy, cross-border data flows, understanding domain name markets in India, impact of internet shutdowns on the economy, intermediary liability, and cybersecurity policy development.

- **Kritika Bhardwaj** is an independent legal practitioner. Her practice involves commercial and regulatory disputes, including in the field of privacy, data protection and telecommunications law. She has also appeared / assisted in pro - bono litigation related to free speech and privacy. Kritika writes frequently on contemporary privacy and data protection concerns.

- **Mansi Kedia** is Senior Fellow at the Indian Council for Research on International Economic Relations. Her areas of research include telecommunication policy, trade and industrial policy. Her ongoing research includes issues related to governance of the Internet, competition issues in India's mobile handset industry, internet shutdowns and the future of work in India.

- **Siddharth Peter de Souza** is a post-doctoral researcher at the Global Data Justice Project at Tilburg Institute for Law, Technology and Society. His work explores how data is governed globally in contested, and plural settings, and he is interested in the role that social movements and civil society can play in shaping governance frameworks. He has recently published a monograph titled Designing Indicators for a Plural Legal World (Cambridge University

Press 2022) which discusses how rule of law indicators need to be reimagined to account for legal pluralism, and contexts and countries in the Global South. Siddharth was previously a PhD researcher at Humboldt University, Berlin, a German Chancellor Fellow, at the Max Planck Foundation for International Peace and the Rule of Law, Heidelberg, and a Judicial Clerk at the High Court of Delhi.

- **Smriti Parsheera** is a lawyer and public policy researcher interested in data governance, privacy, digital competition, and regulatory processes. She is currently a Fellow with the CyberBRICS Project at FGV Law School, Brazil and a PhD candidate at IIT Delhi's School of Public Policy. Before this she was involved in setting up and leading the technology policy work at the National Institute of Public Finance and Policy. Smriti studied law at the National Law School of India University Bangalore and the University of Pennsylvania.

- **Vidushi Marda** is a lawyer and researcher who investigates the societal implications of artificial intelligence (AI). She works at ARTICLE 19, where she leads research and engagement on the human rights implications of machine learning. Her work has been cited by the Supreme Court of India in a landmark ruling on the Right to Privacy, the United Nations Special Rapporteur on freedom of opinion and expression, among others. Vidushi's policy work includes providing input on resolutions at the United Nations Human Rights Council and General Assembly, delivering expert testimony at the European Parliament, and contributing to AI and data policy development in multiple jurisdictions including India, the European Union, United Kingdom and United States of America. She is a member of the Expert Group on Governance of Data and AI at United Nations Global Pulse, and part of the Founding Group of REAL ML.

We appreaciate the time and attention you have invested in this book.
Hope you had a good time!
We welcome your feedback, suggestions, ideas and insights.
You can write to us at **ccg@nludelhi.ac.in**

Centre for Communication Governance at
National Law University, Delhi | Sector-14,
Dwarka, New Delhi - 110078, India
ccgdelhi.org | privacylibrary.ccgnlud.org
Email: ccg@nludelhi.ac.in @CCGNLUD